MODELING AND PREDICTING THE VARIATION OF US HIGHWAY CONSTRUCTION COST

A Dissertation
Presented to
The Academic Faculty

by

Yang Cao

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Building Construction

Georgia Institute of Technology
May 2019

MODELING AND PREDICTING THE VARIATION OF US HIGHWAY CONSTRUCTION COST

Approved by:

Dr. Baabak Ashuri, Advisor
School of Building Construction and School of Civil and Environmental Engineering
Georgia Institute of Technology

Dr. Iris Tien
School of Civil and Environmental Engineering
Georgia Institute of Technology

Dr. Javier Irizarry
School of Building Construction
Georgia Institute of Technology

Dr. Xinyi Song
School of Building Construction
Georgia Institute of Technology

Dr. Mohsen Shahandashti
Department of Civil Engineering
The University of Texas at Arlington

Dr. Shuai Li
Department of Civil and Environmental Engineering
The University of Tennessee Knoxville

Date Approved: February 27, 2019

To my family who always support me

ACKNOWLEDGEMENTS

I spent two years in master research and four years in doctoral research. Six years' graduate life not only improved my academic and research capability, it trained me to be a man from inside: tolerance, endurance and perseverance. Thanks for my parents and wife, Mrs. Liang Deng, who kept on supporting me with patience in this long process. They are my strongest support behind me all the way. Appreciate for my PhD advisor, Dr. Ashuri, who acted as my academic and life mentor. He is an advisor who shows respects to every student and treats students as friends and colleagues. He endowed me freedom to choose the research topic and gave me instructions without any reservation. Appreciate for my master advisor, Dr. Song, who lifted the bridge for me to study in United States. Thanks for my friends who worked with me in this long process, Jun Wang, Jianli Chen. Also appreciate for all committee members who provided me guidance to graduation: Dr. Irizarry, Dr. Shahandashti, Dr. Tien and Dr. Li.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

CCI	Construction Cost Index
HCCI	Highway Construction Cost Index
NHCCI	National Highway Construction Cost Index
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MSE	Mean Square Error
RMSE	Root Mean Square Error
FHWA	Federal Highway Administration
FAST	Fixing America's Surface Transportation
DOT	Department of Transportation
BPI	Bid Price Index
EEMA	Engineering Estimates and Market Analysis
CMA	Contracts and Market Analysis
ENR	Engineering News Record
GDP	Gross Domestic Product
ARIMA	Autoregressive Integrated Moving Average
KNN	K-nearest-neighbor
PERT	Perfect Random Tree Ensembles
LSTM	Long Short-term Memory
Seq2seq	Sequence to sequence model
PAC	Partial Autocorrelation
AC	Autocorrelation
NTP	Notice to Proceed

SUMMARY

The U.S. government attaches great importance to highway construction every year. Because of the importance of highway construction projects and the tremendous expenditure, the budgeting and cost control is a significant job for all federal agencies and state Department of Transportation (DOTs). A major problem for highway construction costs is that they exhibit a significant variation over time such as the highway construction cost indexes (HCCI), and it is mainly caused by a complex interactive effect such as market-related and project-specific factors. The variation hinders the estimators to catch the correct trend of the market and thus poses a challenge for both owners, such as state, local Department of Transportation, and contractors, in correct budgeting and cost estimating.

The main objective of this PhD research is to explore the explanatory variables and develop machine learning methods to model and predict highway construction cost and examine the accuracy of the forecasting results with those of the existing methods such as regression analysis and time series models. The major promising feature of the proposed methods over existing ones is the capability to explain the non-linear relation, which is prevailing in practice. Besides, there is no restriction for the proposed models, compared to linear regression, which assumes the linear relation and normal-distributed-error, and time series analysis, which requires the stationarity of data before analysis.

The dissertation summarizes the results from two parts of the research: (1) Univariate time series index prediction in Long Short-Term Memory; (2) Modeling the

unit price bids submitted for major asphalt line items in Georgia, project-specific and macroeconomic factors.

A deep learning model based on the recurrent neural network will be developed due to its strength in long term memory and catching the variation of the data. The Texas HCCI is used as an illustrative example because Texas reports a frequent volatile index. The performance of the model will be tested in different prediction scenarios (short-term, midterm and long-term). A dataset containing fifty-seven variables with potential power to explain the variability of the submitted unit price bid are collected in this study. These variables represent a wide range of factors in six aspects: project characteristics, project location and its distance to major supply sources for critical materials, level of activities in the local highway construction market, overall construction market conditions, macroeconomic conditions, and oil market conditions. The machine learning feature selection algorithm Boruta analysis will be used to select the most significant features with the greatest capability to predict the unit price bid for asphalt line items. The partial dependence plots endow the explanatory power to the developed machine learning models. An ensemble learning model will be constructed based on the selected features to forecast the unit price bid. The accuracy of the predicted machine learning models will be compared and validated with the existing multiple regression model and the Monte Carlo Simulation.

The main contribution of this research to the body of knowledge in cost forecasting are be summarized from three aspects: first, the research developed a new set of machine learning models that provide more accurate costs forecasts compared to the existing methods. For example, the non-linear machine learning methods are more

accurate than the time series models which are frequently used in the former research. In the field of cost research, a small improvement in model accuracy results in a significant amount of actual impact in budget estimation. Second, the modified encoder and decoder architecture performed well in numerical time series data prediction problem. Instead of making one sequence of output, the roll-forward forecasting turned out to be more accurate. Third, from the practical application perspective, the proposed machine learning models can handle a wide range of issues with the input data that are common in the field of highway construction cost forecasting, such as missing values. Another practicality contribution is that methods in this research are applicable to big data which is an industry trend, while most former models were developed based on a small dataset. The research also proposed a construction cost database which will largely provide the convenience for easy utilization of the model. Fourth, the research identified the most significant features to forecast the variation of unit price bids of resurfacing projects in Georgia, and the analysis laid emphasis on the explanatory power of prediction models. With the improved prediction capability, state DOTs can benefit from the proposed models in preparing more accurate budgets and cost estimates for highway construction projects. The analysis process and proposed models in research are also applicable to other time series data prediction and cost estimating problems.

1. INTRODUCTION

1.1 Background

1.1.1 Highway Construction Cost Index (HCCI) and Its Variation

Highway construction cost index (HCCI) is a compositive price index to reflect the average changes in the prices of the industry over a period of time. It is a unitless indicator composed of price information (material, labor, and equipment) for major line items in the highway construction projects. There are both national level and state level index in practice. National highway construction cost index (NHCCI) was published by federal highway administration (FHWA) as a macro-level index, while over ten states have developed their own state level HCCI, because sometimes the NHCCI could not accurately reflect the local market condition (Huntsman et al. 2018). Based on a national survey from the research of Shrestha, Jeong and Gransberg, the HCCI has four major applications: it could be used as a cost inflation factor; it could be regarded as a general construction market indicator; it could be considered as an index for the purchasing power of the federal or state agencies; and it could also be used to compare the market condition between the whole country and one state, or between any two neighbor states (Shrestha et al. 2016). A major problem for HCCIs is that they exhibit a significant variation over time as other data such as construction cost indexes (CCI) and it is mainly cause by a complex interactive effect such as market-related and project-specific factors (Cao et al. 2018). The variation hinders the users to catch the correct trend of the market and thus poses a challenge for both owners, such as state, local Department of Transportation, and contractors in correct

budgeting and cost estimating. To solve the challenge, a lot of research focused on predicting the HCCIs and explaining the variation using quantitative models.

1.1.2 Highway Construction Cost and the Variation

The U.S. government attaches great importance to highway construction every year. In 2015, former President Barak Obama signed a bill to pass the five-year, $305 billion Fixing America's Surface Transportation (FAST) Act (ARTBA 2016). Current President Donald Trump called on Congress to provide funding of over $1 trillion to upgrade America's roads, airports, and rail lines (Aric 2017). In highway construction, resurfacing is one of the most common projects led by state department of transportations (DOTs) and is used to extend the life of existing highway infrastructure.

"Highway agencies focus more on maintaining and rehabilitating existing roads rather than building new ones, and resurfacing has become their largest expenditure" (Wang and Liu 2012). The need for resurfacing highways is about 24% of all interstate, express, and major highways in the United States, while in some states, such as New Jersey, this number goes up to 35.5%. (ARIBA 2014). "For instance, Virginia DOT (VDOT) spent 49% of its $3.378 billion transportation budget on highway maintenance and operation in 2010, and the budgeted percentage increased to 51% in 2011" (VDOT 2010). Illinois DOT (IDOT) plans to spend over 4 billion dollars, which is about 54% of the total budget on state highways, in reconstruction, resurfacing, widening, and safety projects in 2017 to 2022. Resurfacing is expected to cost 738 million dollars (IDOT 2016).

Because of the importance of resurfacing projects and the tremendous expenditure, the budgeting and cost control is a significant job for all state DOTs when working on

resurfacing projects. Among many indicators, the value of unit price bids is the one that could reflect the comprehensive cost of a unit resurfacing work, which results from the cost of labor, material, machinery, and so on. For each project, the bidding price is mainly related to two kinds of factors: high level economic indexes in the construction market, such as oil price; and project-specific factors, such as the location of the project.

As with many other construction cost indexes, such as the construction cost index (CCI), the unit price bids are undergoing significant variations over time; the comprehensive impact of the two groups of factors already mentioned (i.e., market-related factors and project-specific factors) makes the change of bidding price more complicated. The trend or the volatility of cost index is problematic for cost estimation, bid preparation, and investment planning of capital projects (Ashuri and Shahandashti 2012).

The development of accurate cost estimates is an important part of delivering highway projects. State highway agencies must estimate the cost of projects at several stages of the project development process – from initial planning, through the design phase, and finally, advertisement and award. State DOTs compare the Engineer's Estimate, based on the final design, to the bid prices received from contractors as part of the contract award process. Variance between the Engineer's Estimate and bids can often lead to delays in project award, budget, and project selection problems.

A study conducted by the Construction Financial Management Association revealed that approximately one-third of participant contractors consider the variability in construction costs as one of the most important risks that impact their profits (Ervin, 2007). Moreover, construction cost variations have adverse impacts on public and private owners

of major capital projects (Dayton 2006; Gallagher 2008). It especially poses risk for highway construction projects with fixed-price contracts and contractors' benefit will be impact a lot by price increase and inflation. (Damnjanovic et al. 2009).

An understanding of unit price bids learned from historical data and then, making more accurate prediction by utilizing other indicators help state DOTs to better budget for project costs, as stated in the FHWA report "Top-down construction cost estimating model using an artificial neural network" (Gransberg et al. 2017). More efforts are needed to develop high-resolution forecasting methods to achieve the required double-objective of cost forecasting: high accuracy in prediction and low effort for development and update.

1.2 Research Problems

Even though the cost modeling and prediction problem has been widely explored, many crucial problems have not been solved, and are summarized from the following four aspects.

First, many models can only work well with a less volatile problem and researchers overestimated the performance of the model by solely looking at the prediction accuracy. When researchers were fully focusing on analyzing the error metrics, or sometimes were satisfied with the low error rate, few of them considered the two significant problems: did the dataset (or the problem) need a complex algorithm, and did the algorithm contribute significantly for predicting the data. The example is ENR CCI, there are totally 475 data points from January 1975 to the July 2014. The dataset shows a low volatility, where the change ratio of two consecutive months is calculated by:

$$Change\ ratio = \frac{Y_{i+1}}{Y_i} - 1$$

where Y_i is the CCI in month i. From the histogram the low volatility of the data is further confirmed, because the change ratio is highly concentrated around 0. The average of the absolute value of change ratio is 0.38% based on the calculation, which means the CCI, compared to the last month value, will either increase 0.38% or decrease 0.38% in an average sense.

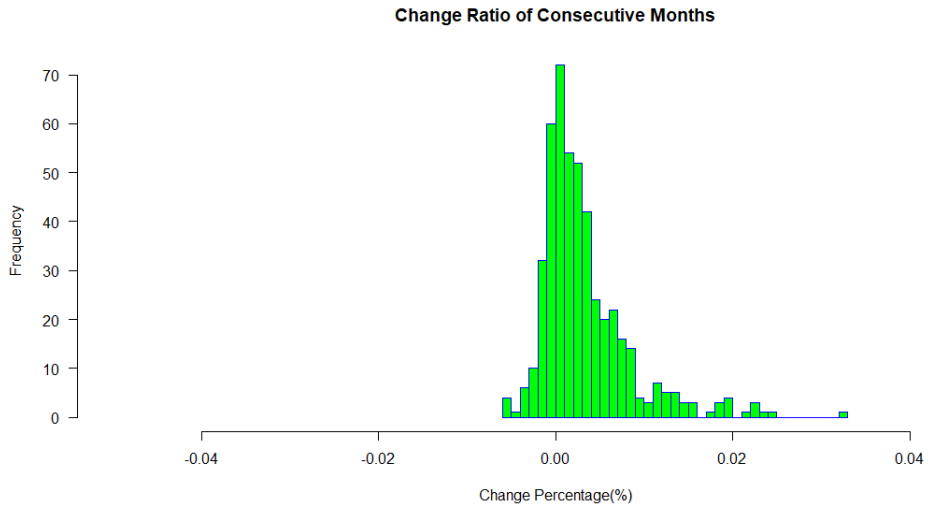


Figure 1-1. Histogram of the change ratio for ENR CCI

This analysis discloses two facts: first, there should exist simple method to predict the data. For example, fitting a straight line or taking the previous month value as the prediction for the value in next month (we define these two models are two of the simple ones). Either model should have a small error and a close-to-one R square. Based on this fact, the second thing is that the error measure might be "inflated". For example, there was

one research predicting the ENR CCI using the time series model and the best out of sample prediction error was about 1% in terms of MAPE. The result looks good at the first glance, but this means that prediction error band is even much larger than the average change ratio (0.38%). In other words, even though the 1% MAPE is small enough but for this problem the model was not good enough for the prediction. Another research with the machine learning model enhanced the accuracy by calculating the out of sample MAPE to 0.18%, then this was a good model only in terms of accuracy (a good model should be evaluated form many perspectives, not just the accuracy). The conclusion is that for the time series data prediction, it is necessary to first analyze the variation of the raw data and have a rough sense of which model might be a good choice and what accuracy level should be achieved using complex models. The existing forecasting methods, such as regression, Monte Carlo simulation, and time series analysis are not satisfactorily robust to predict cost when working with high volatile data. Developing a more robust forecasting method is one of the main objectives of this research to address the complexity of the underlying cost data. The main goal is to enhance the accuracy of prediction under different conditions.

The second problem is there is a need for developing a prediction model that can utilize information embedded in other variables to enhance the prediction of highway construction cost, and developing a method to detect both linear and non-linear explanatory variables of highway construction cost. Several variables, including but are not limited to project-specific features, macroeconomic indicators, indicators of regional and local construction market conditions and representatives of energy (e.g., oil) market conditions are identified as potential leading indicators of construction cost. It is important that the prediction method has the capability to handle a wide range of features that have potentials

to improve cost estimation. Highway cost estimators need to consider the wide range of indicators in an efficient way to select the best subset of features with the greatest potential to predict the cost. There is a need for an efficient forecasting algorithm to select the best subset of features that provide the desired level of accuracy in predicting the cost.

The third problem is about a practical consideration that has not been analyzed before. Considering a wide range of data features is a challenging task. A desirable forecasting model needs to: (a) Handle a large number of variables efficiently and effectively; (b) Work with both numerical and categorical variables; and (c) Deal with missing data points that is unfortunately a common problem in highway construction cost analysis. Therefore, there is a need to develop an efficient forecasting method capable of capturing the complex and (often) non-linear relationships among the underlying variables that can be used to enhance the cost prediction. In addition to addressing the issues related to the nature of input data, the desired forecasting method needs to be scalable to make it useful for practical applications. Transportation agencies invest in ongoing data management efforts. An appropriate forecasting method needs to have the capability to get updated as new features are introduced to the forecasting process and new values of data become available. The required forecasting method should be equipped with an appropriate feature selection technique to examine the relevance of the new variable in the context of all other data features. The forecasting method needs to get updated as time goes by and new values of the data features become revealed.

The last problem is related to the model evaluation. It is a common issue when working with the multivariate time series data. To set up the question, suppose there are three independent time series data X, Y, and Z. The dependent time series O is predicted

from X, Y, Z using a quantitative model. For the purpose of simplicity, the input vector I_i is formed by combining X, Y and Z, so $I_i=[X_i\ Y_i\ Z_i]$, where i means the index of the time. The whole dataset is divided into training set and test set:

$$training: input: I_0 \sim I_t; output: O_0 \sim O_t$$

$$testing: input: I_{t+1} \sim I_T; output: O_{t+1} \sim O_T$$

The model trained through the training set is named as O=f(I), which maps the input I to the output O. Testing could be done through the following way:

$$\hat{O}_{t+1} \sim \hat{O}_T = f(I_{t+1} \sim I_T)$$

The hat symbol means this is the corresponding predicted value. This is a correct testing method except when someone wants to conclude that the model performs well to predict ahead (T-t) period of time. The problem here is that at every testing step the input is automatically updated to the real value! In a strict sense, the model is only tested to predict one point forward. Without the notice of this point, the power of the model will be overstated.

This relates to the discussion that the time series model has a built-in "non cheating" forecasting procedure:

$$training: \ [O_i, I_i] = f([O_{i-1}, I_{i-1}]), \quad i \in (1, t)$$

$$testing: [\hat{O}_{t+1}, \hat{I}_{t+1}] = f([O_t, I_t])$$

$$[\hat{O}_{t+2}, \hat{I}_{t+2}] = f([\hat{O}_{t+1}, \hat{I}_{t+1}])$$

$$\vdots$$

$$[\hat{O}_T, \hat{I}_T] = f([\hat{O}_{T-1}, \hat{I}_{T-1}])$$

From the equations, we can see that the time series model will only use the actual input in the first point of the testing set, and then the output will be used as the next input and so on. This is the strict way that testing does not utilize any "future" information. Necessary attention needs to be paid to the machine learning models, because self-defined testing procedure needs to be created. To conclude this problem, the premise of using the above testing process is to evaluate the capability of a model to predict ahead for a future period of time.

1.3 Research Objectives

Based on the detected problems, this dissertation aims at solving the following two problems, one from the theoretical perspective, one from the engineering aspect: first, exploring methods that is able to deal with the "hard" prediction problems in the industry, in other words, the models are capable in learning the high volatility of the data. Second, there is a need for developing a prediction model that can utilize information embedded in other variables to enhance the prediction of highway construction cost, and the model needs to: (a) Handle a large number of variables efficiently and effectively; (b) Work with both numerical and categorical variables; and (c) Deal with missing data points that is unfortunately a common problem in highway construction cost analysis. Besides, the prediction model should have the potential to be applied in the industry. It needs to be easily implemented with the efficient training time.

1.4 Research Methodology

To fulfill our purpose mentioned above, the research starts with a comprehensive literature review to understand the current practice and the gap of knowledge. Research is further divided into two parts based on the application scenario: either modeling the univariate time series data or modeling the multivariate ones. For the univariate case, the long short-term memory is selected as a suitable model, with an illustrative example using Texas HCCI. The demonstrated model evaluation process proposed in the dissertation is more objective and accurate than former research: it tests the model from long, mid and short-term scenarios and ensures no future information is touched in the testing. For the multivariate case, a significant part of effort was devoted into data collection. This step displays the majority of useful data sources and collection methods in highway construction economics. The application of the non-linear feature selection methodology helps researcher to identify the important features, which could not be realized in Pearson correlation analysis. Feature selection also lays the solid foundation for establishing the ensemble learning model to predict the unit price bids of resurfacing projects in Georgia. The model is applicable to other highway construction industry dataset as well.

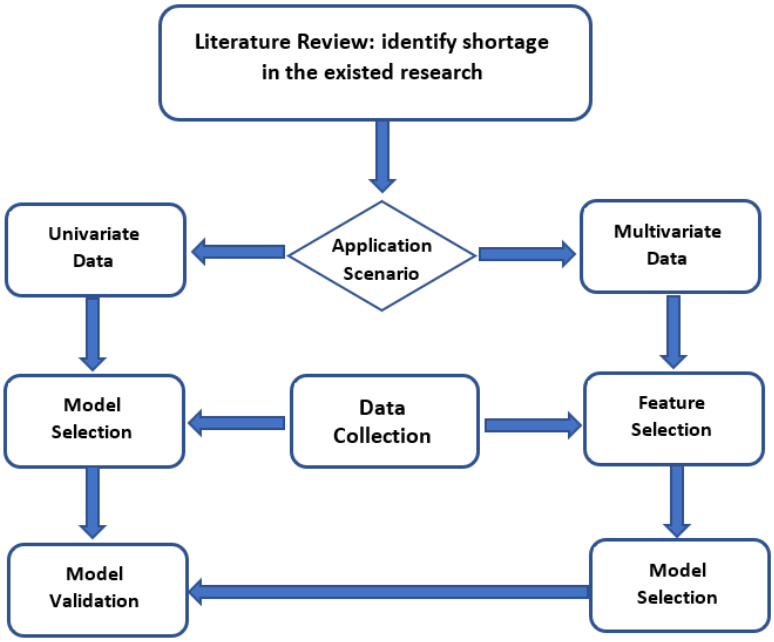


Figure 1-2. Research framework

2. LITERATURE REVIEW

1.1 National and state level HCCI

As mentioned before, the HCCI is a kind of compositive indicator to reflect the price trend of the highway construction industry. The value is mainly determined by major line items related to the material, equipment and labor. The most famous macro-level HCCI is published the Federal Highway Administration (FHWA), and the index is the National Highway Construction Cost Index. Except for this national level index, around 14 states have developed their own index.

Before the NHCCI, the FHWA developed the Bid Price Index (BPI) in order to keep track of the highway construction markets in United States. Until 2003, the BPI was replaced by the NHCCI with a totally different calculation method (FHWA 2014). Since then, it publishes the quarterly index as the price indicator for the highway construction industry for the whole nation. FHWA only uses the winning bid data on highway construction projects, while some states use all the submitted bids for computation. FHWA conducted the calculation based on the Chained Fisher Index, which is regarded as the best practice for the state index development methodology. The base year of NHCCI is 2003 with the base number 1. In 2014, FHWA proposed a major revision after one of its periodic review processes. They found that the price trend was stagnant or even declining near 2014, which was not consistent with the overall inflation level (their benchmark index was producer price index). In a research study in 2015, FHWA made significant effort in

improving the data inputs and estimation problem. After the changes being made, the new NHCCI better reflected the actual inflation level (FHWA, 2017). From the FHWA website, the three changes they made are listed as follows:

"Unit of measure and non-standard pay item issues – The enhancement establishes crosswalk applications that translate inconsistent units and non-standard pay items into consistent units and standard pay items so that these observations can be included in calculating the index. More observations are included in the index calculation because of this improvement.

Changes in statistical exclusion procedure – The new methodology changes the thresholds used for identifying outlier observations, pay items subject to quantity discounts, and observations with extreme price fluctuations. These improvements allow including a wider range of observations, which results in price trends reflective of other national level price indices.

Changes in data reporting by States - States occasionally introduce changes to their pay item numbering system for organizing and reporting construction bid data. More often, when a State changes pay item numbering system, pay item descriptions remain the same, but their corresponding pay item numbers change from one year to another or from one project to another. Such changes in data reporting create a break in the time series of goods included in the index estimation. The revised methodology addresses this issue, enabling the use of more data and more consistent estimates over time." (FHWA 2017)

In Colorado, the office of Engineering Estimates and Market Analysis Unit (EEMA) and the Contracts and Market Analysis Branch (CMA) are responsible for

publishing the CCI. CDOT published the quarterly index and the guidance document was included in the quarterly report in 2012. CDOT uses the fixed weighting factors since it first developed the index, even though they have realized the limitation that it usually overstates the impact of price increases and understates the impact of the price decreases. The good practice of CDOT is it resets the base period every 10 to 15 years, but the last version of index was based on 1987, having been used for 25 years. The calculation method of the index was the same as the old version NHCCI, which was the Fischer Ideal Index. The outlier detection of the CDOT is simple: removes out the data points that falls outside of the 5% and 95% percentile of a given sub-group based on the past seven years data. There are five sub-groups of lime items in the CDOT HCCI: earthwork, hot mix asphalt, concrete pavement, structural concrete and reinforcing steel. There used to be six sub-groups, but the structural steel group was deleted on 2012 because it costs less than 1% of the total investment annually. The current HCCI uses the 2012 as the base period (CMA CDOT 2012).

In California, the office of Construction Contract Awards is responsible for publishing the HCCI and Caltran uses the quarterly index. The current index was developed using 2007 as the base year. Caltran divided the line items into seven sub-groups: roadway excavation, aggregate base, asphalt concrete pavement, Portland cement concrete (pavement), Portland cement concrete (structure), bar reinforcing steel and the structural steel. The very different place in Caltran is that their calculation is based on all bidders' prices for the selected bid items in the corresponding quarter, while most of other DOTs use the data from awarded bidder or actual payment. Caltran also reports the average number of bidders per project (Caltran 2018).

In Washington State, the Construction Office used to publish the HCCI but their last update was in 2016 and they only published the annual data. The calculation was based on seven bid items to track the trend of the state's material price change. The seven selected items were: roadway exaction, crushed surfacing, hot mix asphalt, Portland cement concrete pavement, structural concrete, steel reinforcing bar and the structural steel. The base year was 1990 with the base number as 110 (WSDOT 2016).

In Iowa, the office of contracts in Iowa DOT takes the responsibility to publish the price trend index. The office publishes the quarterly data, but it also calculates the three-quarter and the one-year moving average of the index. The base year for the current index is 1987 and the base index is 100. Iowa DOT computes the index based three construction categories and six indicator items: category roadway exaction is composed of one item, class 10 roadway and borrow, and embankment-in-place. The category surfacing is composed of two items: HWA pavement and shoulder mixes and class a, b, c P. C. C. pavements. The category structure is composed of three items: reinforcing steel, structural steel and the structural concrete. Iowa DOT sets a lower limit for each item in a project to be considered for the index calculation. For instance, only the roadway excavation no less than 22937 cubic meters will be used in calculation. The calculation sample of Iowa DOT is the price from all awarded contracts (Office of Contracts, Iowa DOT 2018).

The Utah DOT publishes two indexes: one is called the construction cost index for roadway excavation, surfacing and structures; and the second index is called construction cost index for all average unit bid prices. The first index is meant to represent the roadway reconstruction and brand-new roadway construction. The second index is to represent all project types for the year. The index calculation is based on six indicator items: roadway

excavation, hot mix asphalt, P. C. C. P. (9'' to 11'' thick), reinforcing steel (coated), structural steel and the structural concrete. The base year of Utah index is 1987 with the 100 as the base number (Utah DOT 2019). The Utah DOT applies the modified Laspeyres methodology to calculate the index (MNDOT 2018).

In Minnesota, the office of project management and technical support from Minnesota DOT takes the lead to develop the state HCCI. Minnesota DOT publishes both composite index and three sub-index every quarter. Six indicator items compose the three sub-indexes: the excavation sub-index is composed of excavation item; concrete pavement and plant-mixed bituminous, form the sub-index in structures, indicating the price trend for all surfacing types; and reinforcing steel, structural steel, and structural concrete, form the price trend for structures. Minnesota DOT first computes the quarterly average unit price for each indicator item, while many other states sum up all the payments. Minnesota DOT excluded all the Design-Build projects and only considers projects costing more than 100,000 dollars. The base year is 1987 with the base number 100 (MNDOT 2017). From another report of Minnesota DOT, "Bituminous pavement makes up 43 percent of the composite index, concrete applications account for a total of 31 percent of the composite HCCI (with a split of one-third for pavements and two-thirds for structures), roadway excavation accounts for 14 percent of the HCCI, and reinforcing and structural steel account for 11 percent." (MNDOT 2018)

In Ohio, the office of estimating of Ohio DOT is responsible to report the state construction cost index. They adopted the modified Laspeyres index calculation method but changed the method to the Chained Fisher index based on a presentation in year 2013. Ohio DOT reports the data monthly. The old base year was set as fiscal year 2004 to 2005

while the new base is the fourth quarter in 2012. Except for the methodology adjustment, two other main changes are increasing of the items and the systematic handling of the outliers. There are total 20 item classes used for forming the index: aggregate base, asphalt, barriers, bridge painting, curbing, drainage, earthwork, erosion control, guardrail, landscaping, lighting, maintenance of traffic, pavement markings, pavement repair, PCC pavement, removal, signalization, structures, traffic control, and unclassified construction (other) (MNDOT 2018). Before the method adjustment, there was only 10 items and with the fixed weights. This change makes the total dollars represented by the selected items increased from 28% to 45% of the total dollars awarded. The second change is Ohio DOT has a better practice in dealing with outliers: the outlier is defined as contract price greater than two median absolute deviations from the median in a quarter. They replace the outlier with the median value (Office of Estimating ODOT, 2013). Even though the presentation did not mention the benefit of outlier processing, it could make the index less volatile and reflect the price trend more accurate.

In Montana, Jeong and his research group proposed a new methodology to calculate the HCCI for Montana DOT. The old practice of Montana was using 52 items from 9 categories to calculate the HCCI and the base year was 1987. The calculation MDOT used was the Young index. However, Montana DOT updated their base year several times in order to get the more accurate index. In 1995, 1997, 2000, 2003 and 2006, the base year was updated because there were significant changes in the item basket. The problem pointed by Jeong was that the base number was not reset to 100 every time (Jeong et al. 2017). The proposed new index will be based on 31 item classes, 10 item types and 6 item divisions: general provisions, earthwork, aggregate surfacing and base courses, bituminous

pavements, rigid pavements and structures, and miscellaneous construction. The calculation method will be changed to Chained Fisher Index. Two innovative points suggested by Jeong and his colleagues were calculating the multidimensional index based on the project size, type and location; and using a dynamic item basket. Instead of keeping the same item basket, it should be updated frequently based on the actual purchase behavior of DOT (MDOT 2018).

In Texas, DOT first published the HCCI in 1998 and the index tracked the price and quantity of 34 highway construction control items, such as lime, and lime treatment. The 34 control items further compose 16 elements, such as lime treated subgrade or base. These elements finally compose the four categories: earthwork, subgrade and base course, surfacing and structures. One of the good practices of Texas DOT is that it publishes the above four category index monthly as well, except for the HCCI. The sub-indexes help the DOT to better understand the price trend in each work type. Another good practice is that Texas DOT not only published the monthly index, it also reports the 3-month moving average and the 12-month moving average. A longer window helps to reduce the volatility and exhibit more clearly on the long-term trend.

1.2 Quantitative Modeling in Highway Construction Cost/Index

Prediction and estimation of highway and construction cost index makes up an important area in highway construction research. Various quantitative models have been applied to predict different cost indexes. Most research aims at predicting national- or state-level cost indexes, such as the construction cost index, national highway construction cost index (NHCCI), and asphalt cement index, while some researchers attempted to estimate

the highway project bidding price. Based on the quantitative methods applied in the previous paper, four major types of research could be summarized from this work.

Some researchers defined the linear relation between predictors and response cost index. Minsoo applied the stepwise regression analysis to select the related factors to explain the variation of the unit price bids for resurfacing projects in Georgia (Baek 2018). Lowe and his colleagues constructed the linear model between 41 independent variables and the construction cost of buildings. The best regression model could give a mean absolute percentage error of 19.3% (Lowe et al. 2006). Instead of forecasting the exact figure of response variable, Ng and his colleagues built the model to predict the change direction (either up or down) of Hong Kong tender price index with eight local indicators, such as gross domestic product (GDP) and unemployment rate (Thomas et al. 2000). Similarly, Wang and Liu also turned to some overall relationship forecasting: they focused on the average bidding price of resurfacing projects in Kentucky to see the relation of bidding price to four factors, including local asphalt price index, because the predominant type of highway pavements is asphalt concrete pavement (U.S. Department of Transportation 2007). Even though a linear regression model was established, the main idea of the Wang and Liu was to know the high-level relationship (positive or negative correlation) between the average bidding price and other indicators. Except for the application of the classical linear regression model, Hwang (2009) applied the dynamic regression, which considered the autoregressive property of the data, to set up the relation between response, CCI, and three predictors: prime rate, housing starts, and consumer price index. Hwang's model was proved to be more advantageous than the existing ones, such as linear regression. All the above work established the direct linear relationship between

predictors and various cost index. These models were restricted to the simple linear relation and neglected the nonlinear relationship, which is not practical to deal with different kinds of data.

Time series analysis is another prevailing method to model and predict many different cost indexes. Univariate time series makes the index prediction based on its own historical records. Ashuri and Lu (2010) conducted the traditional time series analysis on CCI. Four time series models were tested and compared concerning in-sample and out-of-sample prediction accuracy. All models work well in CCI data in terms of prediction accuracy, but that is due to the long-term stationary form of the data. Proposed models do not perform well when some big volatility happens (Ashuri and Lu 2010). Besides, the univariate time series models do not have explanatory capability and they are just suitable for short-term forecasting (Goh and Teo 2000). Moon and his colleagues applied the autoregressive fractionally integrated moving average model to predict the Engineering News Record (ENR) construction cost index, and the model outperformed the benchmark autoregressive integrated moving average model (ARIMA) with the mean absolute percentage error about 9.5 percent. Their model has the advantage that the data does not need to be Gaussian distributed (Moon et al. 2017). Joukar and Nahmens emphasized the significance of modeling the variation of the ENR CCI and the method they chose was the general autoregressive conditional heteroskedastic, which established the connection between volatility and residuals instead of the original index. The out of sample R^2 was over 60% which means 60% of the variation could be explained by the model (Joukar and Nahmens 2015).

The other kind of research in time series is the causal method, which uses the explanatory indicators to structure the relationships between predictors and predicted variable. Mohsen, who did a lot of work in this category, established the relationships between many macroeconomics indicators and CCI and NHCCI. The Pearson correlation test and Granger causality test are used to find the correlation and lead-lag relationship, respectively. Mohsen found that consumer price index, producer price index, GDP, and money supply are useful to predict CCI, while crude oil price and average hourly earnings are the leading indicators of NHCCI (Ashuri and Shahandashti 2012; Shahandashti 2014). These works laid the foundation for predicting response variables through modeling time series on other leading indicators. This is an improvement from linear prediction models as Mohsen argued that the rejection of significant linear correlation does not mean that crude oil price and average hourly earnings do not contain information that is helpful to predict NHCCI (Shahandashti and Ashuri 2015). Mohsen applied the multivariate time series to predict NHCCI with average hourly earnings and crude oil price index. Again, the good performance of the multivariate time series model is limited to short-term forecasting, and the model is not good at catching large variation.

Besides the time series model, which is a kind of a stochastic model, another kind of stochastic model was built by Mohammad Ilbeigi (2014), who applied the geometric Brownian motion (GBM) to model the fluctuations of asphalt cement price in Georgia from 2005 to 2012. The dataset fit well with the form of GBM, but the violation of independence assumption of GBM affects the prediction accuracy of the model. It could be found that the large confidence interval challenges the practical value of this prediction.

Machine learning algorithms are the more advanced alternative to make the cost prediction. They can learn from input variables and make data-driven predictions on output variables, rather than following strictly static prediction models, such as those found in time series analysis methods (Bishop 2006; Wang 2014).

With the help of other explanatory variables (always called *features* in the machine learning scenario), the machine learning model exhibits a better capability of catching the variation and provides the more accurate prediction. Ashuri and Lu (2010) applied the time series model to predict the CCI, and most time series models achieved a good accuracy, measured by mean absolute percentage error (MAPE), to around 1%. Wang and Ashuri (2016) further improved this accuracy to around 0.8%, by applying the machine learning algorithm's k-nearest-neighbor (KNN) and perfect random tree ensembles (PERT) to forecast CCI by consumer price index (CPI) and gross domestic product (GDP). Compared to the research of former time series model, the prediction accuracy has been significantly increased with the extra information from CPI and GDP (Wang and Ashuri 2016). The relatively stationary form of CCI data and the high correlation of the variables restrict the achievement of the model to other dataset. The accuracy of the above model highly depends on the form of data and might be not suitable for other highway cost indexes. Cao and his colleagues designed an ensemble learning model which was composed of four machine learning algorithms to predict the unit price bids in Georgia. The random forest algorithm helped to select most significant 20 features among the 57 collected feature series. The model was also applicable in other data such as HCCIs (Cao et al. 2018).

Neural networking is famous for its capability to model the complex nonlinear relationship among predictors and response. Williams (1994) developed a back-

propagation neural network model to predict short-term changes in the CCI using several variables, such as prime lending rate, housing starts, and the month of the year. Wilmot and Cheng fitted an artificial neural network model to predict Louisiana highway construction cost index with sub-item costs, including embankment, concrete pavement, asphalt pavement, and so on. The results showed that the model is able to simulate the historical change and predict the future cost index with reasonable accuracy (Wilmot and Cheng 2003). Gransberg and his colleagues proposed a top-down cost estimating method using neural networks to improve the accuracy of preliminary cost estimation for Montana DOT, the input variables included the design related variables such as volumes of excavation and embankment; roadway information such as site topography; and construction administration attributes such as letting date (Gransberg et al. 2017).

In addition to the above well-established models, there are multiple simple methods summarized from industry practice. Anderson and colleagues pointed out a simple escalation approach to estimate the cost of highway projects. Cost estimators inflate the estimated cost of materials to the expected midpoint of construction date in order to capture possible changes in the future prices of materials in their estimates (Anderson et al. 2006; Ilbeigi et al. 2016). Another probabilistic-based method was mentioned by Back and colleagues, who used the Monte Carlo simulation to quantify the range and the likelihood of the project cost (Back et al. 2000). At any point in time, Monte Carlo simulation randomly generates an escalation rate independent from the previous rates. Ilbeigi argued the disadvantage of the Monte Carlo simulation estimation method since it ignores the effects of autocorrelation of the historical data (Ilbeigi et al. 2016).

These methods are easily implemented, but they produce an extensive error, especially when estimating a micro-level highway bidding price. Compared to national level data, such as the national highway cost index, which reflect the macro-level information, bidding price data come from a state or a city. “Local costs for items such as labor, steel, or oil can flux widely and have volatile swings that are not always captured in national indexes. Many states use local data and calculate an inflation rate that reflects current conditions” (Thomas et al. 2000). Based on the authors’ empirical research, the local average unit price bids for resurfacing projects has relatively high volatility and the change trends are difficult to catch.

3. PREDICTING HIGHWAY CONSTRUCTION COST INDEX IN LONG SHORT-TERM MEMORY

3.1 Introduction

The research in this part focuses on a univariate time series problem, and the research clearly defines the difficulty of the prediction problem. The Texas HCCI is such a hard prediction problem that is well modeled in Long Short-term Memory. The model validation process is based on three prediction scenarios: long-term, midterm and the short-term forecasting. The self-implemented rolling forward validation procedure ensures that no future information is included in testing process, and thus strengths the validity of model evaluation.

3.2 Research Objective

The major objective of this research is to explore a suitable method to solve the hard prediction problems. It turned out that the advanced deep learning algorithm, long short-term memory (LSTM), successfully overcome the challenges with its powerful structure and the authors-defined forecasting procedure. More specifically, the sequence to sequence (seq2seq) architecture of LSTM will be utilized. The seasonal ARIMA model will be selected as the benchmark model for three reasons: first, it is one of the most widely applied forecasting models based on the literature review, and its effectiveness has been verified by a lot of researchers. Second, the seasonal ARIMA model is a typical linear model and therefore the comparison between the LSTM and seasonal ARIMA gives a better understanding of how linear and non-linear models perform differently; last but not

the least, seasonal ARIMA model has a typical, built-in “non-cheating” forecasting procedure as discussed above.

Different training processes of LSTM and seasonal ARIMA will be displayed and the forecasting performance will be evaluated based on three scenarios: short term, midterm and long-term prediction. The “non-cheating” forecasting method in three scenarios strengthened the conclusion that the LSTM is a powerful and promising time series index prediction tool.

3.3 Research Methodology

3.3.1 Research Framework

The research is composed of three main steps: data collection, model development and the model validation. The first step will also include the data split to fit for the cross-validation in three scenarios. The model validation will use the MAPE as the error metric.

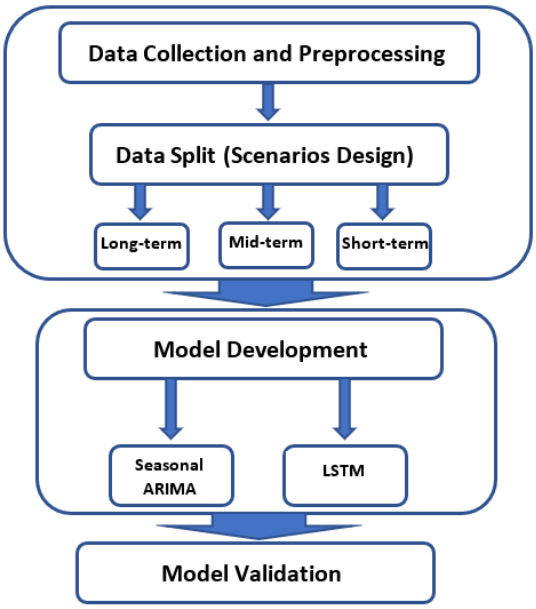


Figure 3-1. Research framework for the first-part

3.3.2 Data Collection

In this research, authors chose the HCCI from Texas DOT and selected the old version, which is 1997-base index and ranging from 1998 to 2017. The new published index is 2012-base and only ranging from 2013 to 2018. The rationale of the selection is from the three-folds consideration:

First, Texas HCCI is one of the few indexes that published monthly, and the partial reason to choose the old version is that the relatively large sample size would be better suitable for training a complex model. In other words, the complex model is supposed to exhibit the more obvious advantages dealing with bigger data.

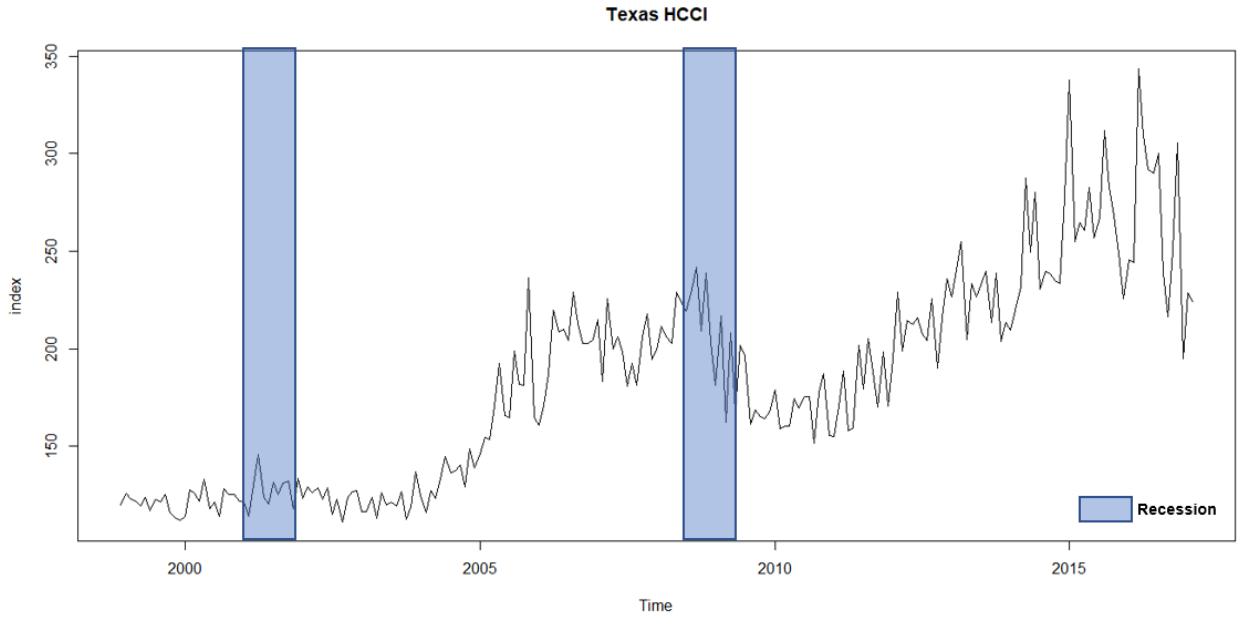


Figure 3-2. Texas HCCI highlighted with the recession periods

Second, based on the literature review, Texas DOT is one of the state agencies that produce the effective and authoritative indicators. Except for the HCCI, Texas DOT also publishes the multidimensional indexes in earthwork, structure, subgrade, and surfacing. Four sub-categories are further based on roadway embankment, cement, surface treatment aggregate, hot mix asphalt concrete, bridge rail and bridge slab. The lower level index could better explain the variation in each sub-category and thus is recommended for DOTs (Jeong et al. 2017). The effectiveness of the index could be reflected in its ability to explain the economy of the state. Historically there were two recession periods in Texas around 2001 and 2008. The economy of Texas started to revive around 2003 due to the strategic development of the high-technology industry. At the same time, it could be observed that the HCCI also started to increase. In the period of well-known worldwide crisis around 2008, the HCCI declined after being volatile for about 3 years. In 2010, the HCCI bounced up along with the recovery of the state economy.

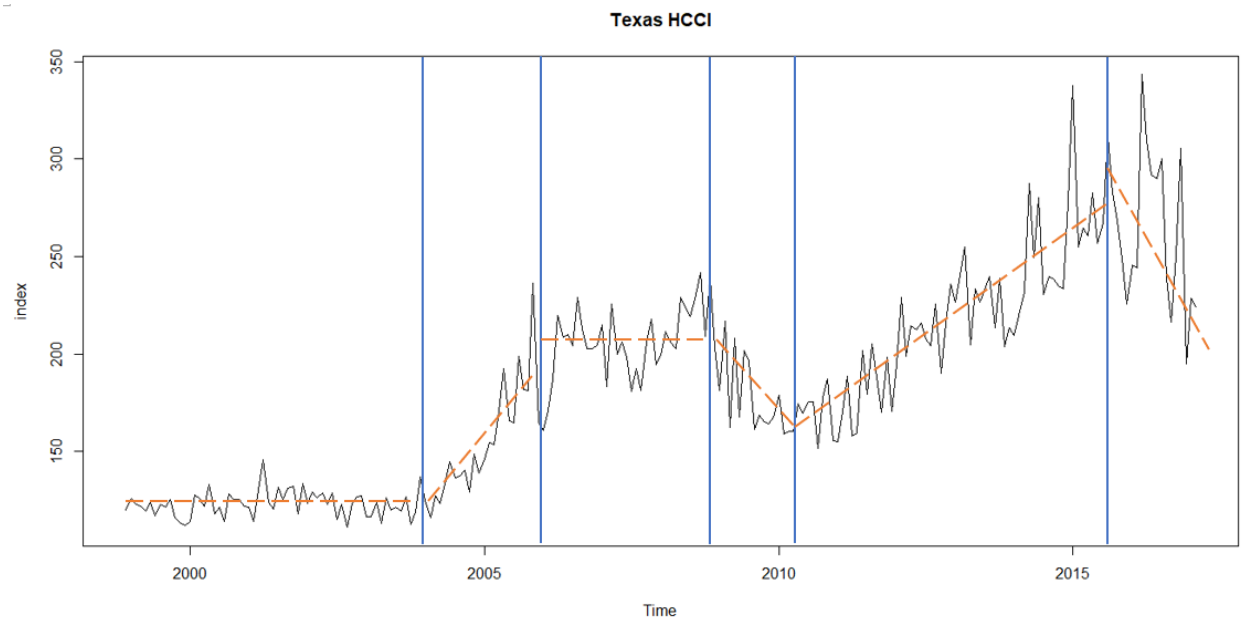


Figure 3-3. Texas HCCI marked with the trend change

Third, this is a "hard" prediction problem. As described in the last point, the obvious trend pattern change could be observed from the index: stable period from 1999 to 2004, followed by a growing period lasting for 2 years. Before the 2008 crisis there witnessed a three-year volatile period. After declined for 2 years, the index increased significantly for five years and declined for two years. This complex pattern brings a challenge to the prediction model. Furthermore, the analysis of the change ratio could be done as discussed in the last section. The change ratio was defined by the percentage change of the index between two consecutive months. Compared to the example for ENR CCI, Texas HCCI undergoes a much higher volatility: the change ratio could be high to 30%. The calculated absolute value of change ratio, which is the average volatility, is 8.45%, over 20 times higher than ENR CCI (this could also be supported by the figure that the change ratio has fatter tails on both sides compared to ENR CCI). In a word, the complex pattern and the

high volatility of the Texas HCCI make it to be a very typical example to demonstrate the power of seq2seq architecture.

3.3.3 Long Short-Term Memory

As of the time author finished the draft, few research used the LSTM in solving prediction problems of construction industry, while the method is a very promising and powerful prediction tool. As suggested by the name, LSTM has the prominent performance in long time memory. It is a special kind of recurrent neural network, which is composed of a more complex unit structure compared to the vanilla neural network. It is first proposed by scientists Hochreiter and Schmidhuber in 1997. The structure ingeniously solves the vanishing gradient problem, which is a commonly issue in the gradient descent process but difficult to be detected when training a neural network (Sundermeyer et al. 2010). The solution is shown in the below unit structure, which has three important parts: input gate, output gate and the forget gate. Both the cell state and hidden state are calculated by the three gates but there is a direct path for the cell state to be able to go from the start all the way to the end, without the gradient update every step, and thus solves the vanishing gradient problem. From an abstract perspective, this mechanism helps to restore the long-term memory.

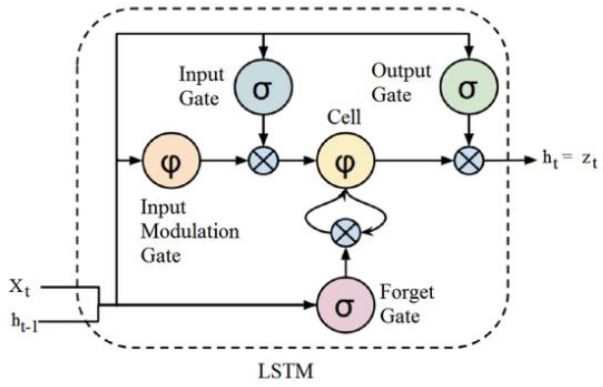


Figure 3-4. The structure of the LSTM unit.

(Figure from https://medium.com/@kangeugine/long-short-term-memory-lstm-concept-cb3283934359)

3.3.4 Encoder and Decoder Architecture

The invention of the LSTM unit dramatically changed the machine learning community, and the subsequent research found that different structure displayed diverse performance on specific problems. The encoder and decoder architecture, also named as the sequence to sequence model (seq2seq), is especially useful in this research. When first proposed, it was a special architecture for LSTM that exhibited a superiority in natural language processing. As demonstrated in the figure, it maps a sequence of input data (encoder) to another sequence of output data (decoder) through a connection named as the intermediate state (hidden state and cell state) (Eddy 2018). Three Google engineers first proposed this architecture for LSTM and tested it in a language translation problem (Sutskever et al. 2014). A lot of following research expand the application field to numerical time series data, and it is one of most popular deep learning structures now. Wang mentioned in his blog that there is an amazing effectiveness of the seq2seq structure for time series data prediction. The structure is flexible for dealing with the variable input

and output sequence lengths. He used a multivariate example to predict the air pollution level and the structure could also be applied in predicting extreme events and outliers (Wang 2017). Another advantage of the seq2seq when compared to the autoregressive integrated moving average (ARIMA) is the ability to handle the high dimensional data (Eddy 2018). This research also applied the seq2seq model to make an accurate prediction.

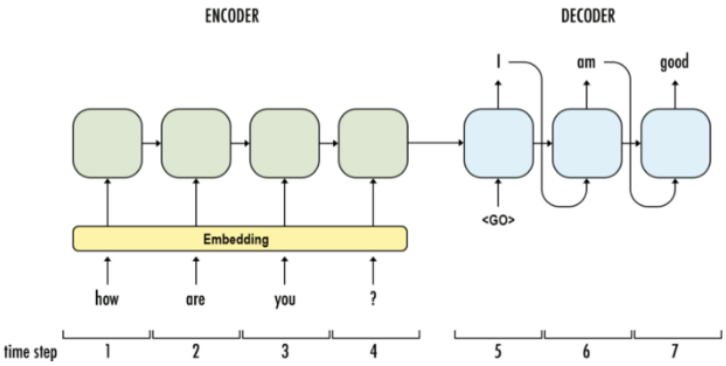


Figure 3-5. The encoder and decoder architecture

(Figure from https://towardsdatascience.com/sequence-to-sequence-model-introduction-and-concepts-44d9b41cd42d)

3.3.5 LSTM in Construction

There is limited application of LSTM in construction industry as it is a relative new algorithm, compared to other machine learning techniques. However, the strong prediction capability will draw attention from more researchers. In the current years, some researchers have attempted to apply the technique in transportation and construction industry and received the positive results. Li and the colleagues applied the LSTM in predicting the traffic state identification. The algorithm helped researchers to achieve a high accuracy by incorporating the imputation data processing method (Li et al. 2018). Zhang and his

colleagues applied the LSTM to predict the multi-step travel time on urban expressways. They compared the algorithm with instantaneous and Naïve k-nearest neighbors algorithms, and the LSTM performed better at this problem in terms of out of sample accuracy (Zhang et al. 2018). Several other application scenarios are found in the safety research. Ding and his colleagues used the LSTM to detect the unsafe construction behavior using computer vision technic. The classification problem defined by if there was any unsafe behavior was tested using concurrent neural network and the LSTM. The experiment supported the effectiveness of the algorithm (Ding et al. 2018). A related research from the same group specifically worked on the fall problem in the construction site and the concurrent neural network was applied to detect the risk, and the precision was high to 99%. The potential of the LSTM in providing an accurate prediction especially in time series data, compared to other statistical methods will result in a wider application of the technique.

3.3.6 Model Development

Dataset was split into training set, development set and the testing set. The first training set was determined, which was composed of data from 1998 to 2008 for two reasons: first, the algorithms cannot work with a smaller training set which would result in overfitting using the LSTM. This sample size helped the model to achieve a good accuracy with the minimum data based on the experiment. Second, the separation point was set on the year 2008 intentionally to test the model performance facing with the pattern change of the data.

Training set was used for training the model. Development set was applied for hyperparameters tuning, and this is especially significant for the LSTM model. For the seasonal ARIMA model there are some techniques to determine the hyperparameters beforehand, even though the tuning process helped to select the better parameter sets. Development set was about the 10% of the training data and the ten folds cross-validation was applied.

Any time series data is composed of three components and it could be expressed in the following format:

$$y_t = S_t + T_t + E_t$$

where the S_t is the seasonal component, T_t is the trend cycle component and the E_t is the error term. The analysis of the Texas HCCI is shown in the following figure. From the top to down are original index, trend component, seasonal component and random error respectively. This decomposition algorithm considers the frequency analysis by taking out the seasonal component. One thing needs to be noticed is that the unit is different for each component. For example, the magnitude of the seasonal component is around 4, compared to the trend component around 100, which means that the seasonality effect is weak in the data. However, we could always include all components in the model to increase the accuracy. The decomposition gave an intuitive suggestion on how to determine some of the parameters.

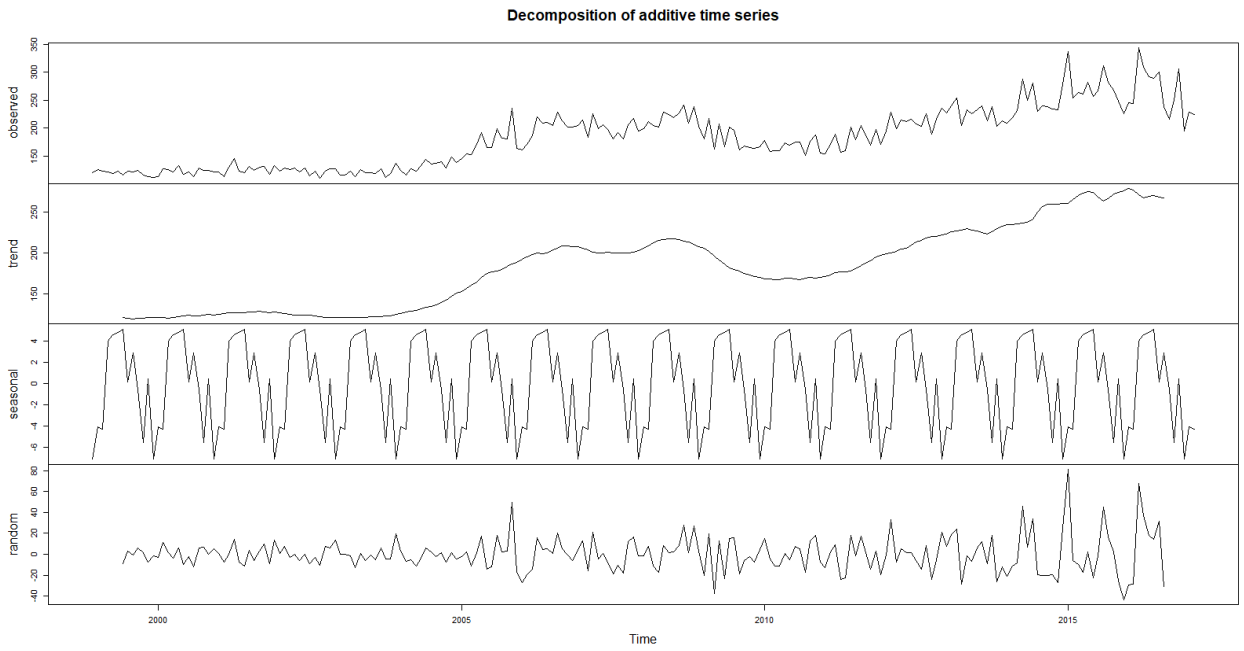


Figure 3-6. Decomposition result of the Texas HCCI

There are typically four steps to train and test a time series model as shown in the figure. The first step is the preprocessing procedure, in order to remove the trend and seasonal component and the model is then applied to fit the error term. The common techniques include the taking logarithm and differencing the original data. Model identification and estimation is to determine the parameters of a time series model such as the order for moving average (MA), order for autoregressive (AR) process and seasonality. Diagnostic checking is the validation on the development set to give a hint that if the parameters need to be changed. After all these three steps finished, the model is ready for future prediction.

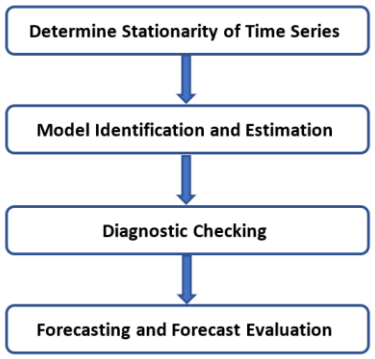


Figure 3-7. Procedures for time series model training and testing

Only three steps for the LSTM modeling but the second step takes much time to form a good model. No requirement such as stationarity is required for LSTM, but the normalization or standardization is recommended as a method to expedite the convergence. In this research the "min-max-scaler" standardizer gathered all data within a small range. This produced a more regular searching space and thus accelerate the convergence of the weights. The second preprocessing step was transforming the univariate time series into the supervised learning. A flexible function was written to be able to transfer the data in any size based on two parameters: input sequence length and the output sequence length. The most important and time-consuming step is the hyperparameter tuning. The method was the variable controlled experiments. The default setting of the model was given in the table. The exact definition of each hyperparameter will be discussed in the next section. When tuning one hyperparameter, others will be set to the defaulted value.

Table 3-1 The default settings of hyperparameters

Hyperparameters	Default Value
Neurons	10
Batch Size	1
Time Steps	12 for input, 1 for output
Epochs	10

The implementation of seq2seq architecture is as follows: first train the encoder LSTM with the input sequence, and then omit all the intermediate output but only save the cell state and the hidden state; second use these two states to initialize the decoder and then fit the corresponding output sequence.

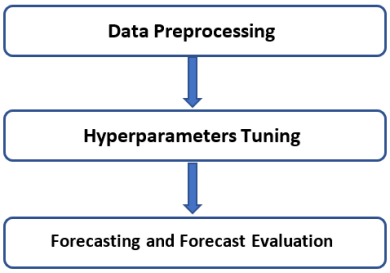


Figure 3-8. Procedures for LSTM model training and testing

3.4 Results and Discussion

The research results are displayed in three scenarios along with the parameters and hyper-parameters selection. Only in the first scenario (long-term prediction), the parameters and the hyper-parameters tuning process and the results will be displayed. The process was the same for other two cases.

3.4.1 Long-term Prediction

The simplest scenario is the long-term prediction, where there is a one-time training and testing. The split of the dataset is shown in the figure. Training methods are very similar and standard for both time series model and LSTM. Detailed discussion is only given for this scenario.

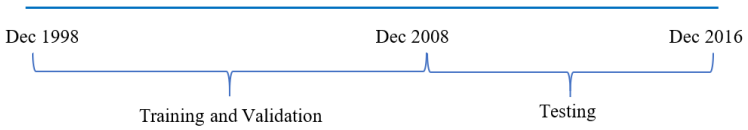


Figure 3-9. Split of the training and testing data

The most common method to determine the MA and the AR term is plotting the autocorrelation (AC) and the partial autocorrelation (PAC) figures. The AC and PAC plots are shown in the figure.

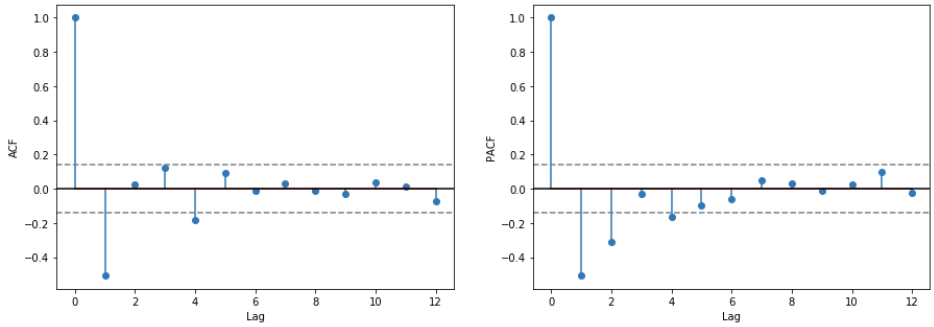
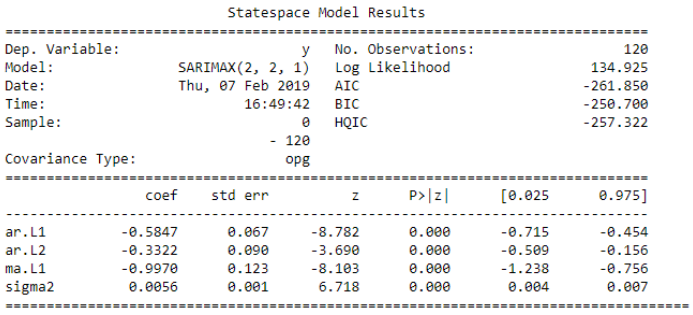


Figure 3-10. ACF and PACF plots of processed index

From the AC figure, only significantly correlated term was the first lag, which was the selection of the MA order. From the PAC figure, first and the second lag were correlated so the AR order was two. From the decomposition figure, the seasonality was determined as 12. In the model training process, the results showed that the seasonality term is not statistically significant as expected from the decomposition figure. The final model was an ARIMA (2,2,1). The statistics of the in-sample fitting are shown in the figure. All the coefficients are statistically significant, and the Bayesian information criterion are the smallest one in in comparison with other nearby parameter sets.



Statespace Model Results

Dep. Variable:	y	No. Observations:	120
Model:	SARIMAX(2, 2, 1)	Log Likelihood	134.925
Date:	Thu, 07 Feb 2019	AIC	-261.850
Time:	16:49:42	BIC	-250.700
Sample:	0	HQIC	-257.322
	- 120		
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
ar.L1	-0.5847	0.067	-8.782	0.000	-0.715	-0.454
ar.L2	-0.3322	0.090	-3.690	0.000	-0.509	-0.156
ma.L1	-0.9970	0.123	-8.103	0.000	-1.238	-0.756
sigma2	0.0056	0.001	6.718	0.000	0.004	0.007

Figure 3-11. The fitting statistics for seasonal ARIMA model

In this research, six hyperparameters were considered: number of neurons, epochs, timesteps, batch size, regularization and the dropout and their results are shown below.

i. Neurons

As this dataset is not too big so there is only one layer. More neurons result in a more complicated model and thus tend to be overfitting. In the LSTM world, another consideration is the training time, and is also crucial for the model to be applied in the highway construction industry. Many experiments were done by trying different number of neurons. As the 10 folds cross-validation was applied, each experiment calculated ten results and their average value was recorded. The error measure was the MAPE. In this step none of the testing set was touched and the usage of validation set was for choosing the optimal hyperparameters.

Table 3-2. Experiment results for tuning the optimal number of neurons

Number of Neurons	Training Error	Validation Error	Training Time
5	0.0632	0.0768	7.5
10	0.0613	0.0669	7.9
20	0.0579	0.0673	8.2
40	0.0552	0.0658	10.6
45	**0.0512**	**0.0614**	**8.45**
50	0.0479	0.0734	7.51

It was inferred from the experiments that the optimal number of neurons was around 40 to 50. The randomness of the LSTM optimization process caused the fact that the training error was not monotonically decreasing as the number of neurons increased. The training time has no relation to the number neurons. The guess was that the model was small so the effect from increasing the neurons is weaker than the randomness. The number of neurons was decided as the 45.

ii. Epochs

The training process of the LSTM is a forward and backward procedure to update the coefficients of the model. One epoch is one round of coefficients update. The effect of the number of epochs on underfitting and overfitting is significant and in academia the so called early-stopping technique was invented: first set a large epoch number and then let the algorithm to find the optimal time to stop training. Here the data is small so experiments were enough for tuning this hyperparameter.

From the result table, authors inferred that the optimal number of training epochs located in the range 150 to 200. The model became overfitting when the training epochs was more than 200. The training time exhibited a monotonic relation to the number of epochs and the slope was big. The number of epochs was set as 175.

Table 3-3. Experiment results for tuning the optimal number of epochs

Number of Epochs	Training Error	Validation Error	Training Time
10	0.0528	0.0627	7.14
50	0.0341	0.0628	11.71
100	0.0242	0.0537	18.56
150	0.0269	0.0496	26.32
200	0.0258	0.0522	33.91
250	0.0255	0.0603	42.14

iii. Timesteps

This is another essential difference between the time series model the seq2seq model. Time series model only predicts one step ahead, at one time, while the seq2seq could predict any length of sequence ahead based on the application. This is not the contradiction with the "non-cheating" forecasting because if the output length is short than the testing period, the predicted value will be used for the following prediction. Authors only tried limited number of combinations because of the business intuition behind the data. For example, it is unrealistic that the state agency does not update the model for many years, and most of the highway projects are finished within two years. The output length means the one-time prediction range. From the table, there was no pattern based on the input-output combination. The guess was because only limited number of combinations were tested. No relation was found between timesteps and the training time. The combination with the smallest validation error was the final choice (24 to 12).

Table 3-4. Experiment results for tuning the optimal timesteps

Input Length	Output Length	Training Error	Validation Error	Training Time
12 (1 year)	6 (0.5 year)	0.0628	0.0845	8.48
12 (1 year)	12 (1 year)	0.0739	0.0917	7.34
24 (2 years)	**12 (1 year)**	**0.0647**	**0.0814**	**7.22**
24 (2 years)	24 (2 years)	0.0725	0.1027	7.86

iv. Batch size

In LSTM the training samples are considered sequentially, and the batch size is like the number of samples being considered at one time. The batch size with one means the

training process uses each individual training sample and it is named as the stochastic gradient descent in the process of updating parameter. A bigger batch size algorithm is so called mini-batch or batch gradient descent optimization, which means a certain amount of training data are used to update the parameters one time. This is an effective way to overcome overfitting by reducing the effect from individual data. In this research, authors chose the batch size as one because the data size is small, the LSTM unit can only receive the number of samples that can be divided by the batch size, so a bigger-than-one batch size means the training sample needs extra change sometimes, in most cases some data needs to be deleted.

v. Regularization and Dropout

Two specific efforts made on the parameters to solve the overfitting: regularization and dropout. For the regularization, even though no regularization term was applied after analysis, the consideration is always recommended. This research first attempted the L-2 regularization terms in training. A good model should not only be dependent on several parameters but in contrary homogenize the contribution of each parameter. A large parameter is a potential sign of the overfitting and thus by adding a square sum of the estimated parameters term in the loss function, the optimization will try to avoid training large parameters. From the results of the experiments, the L-2 regularization term could not make much difference on the accuracy, but the training time took a little longer, so finally no regularization was included in the model. Another similar idea to homogenize the effect of each parameter is the dropout. In the process of updating coefficients in each training epoch, only a portion of the neurons is activated in each layer, controlled by a Bernoulli distribution respectively. This technique will train a neural network not relying

too much on several neurons, instead averaging the contribution of all neurons. An important engineering detail was from the research of Zaremba and his colleagues. They claimed that the dropout should not be performed in the LSTM unit, but only in the regular neurons instead. The dropout in LSTM unit will impair the effect of long-term memory (Zaremba et al. 2014). In this research, therefore, the dropout was only applied in the output layer which was a dense layer in Keras. The ratio was set as 50% recommended by Baldi and Sadowski who conducted simulation to find out the optimal drop out ratio (Baldi and Sadowski 2013).

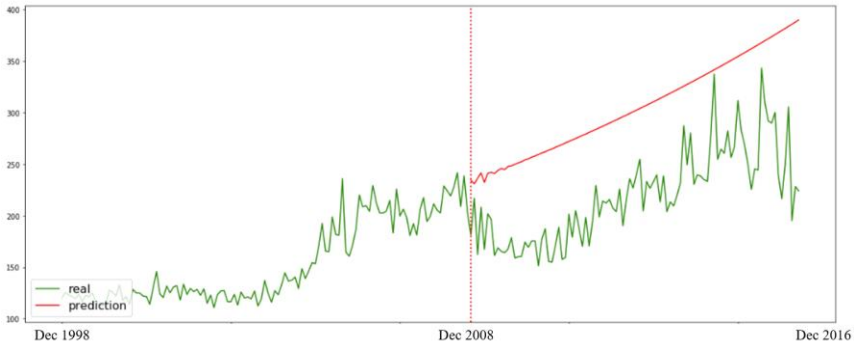


Figure 3-12. Eight years forecasting by seasonal ARIMA model

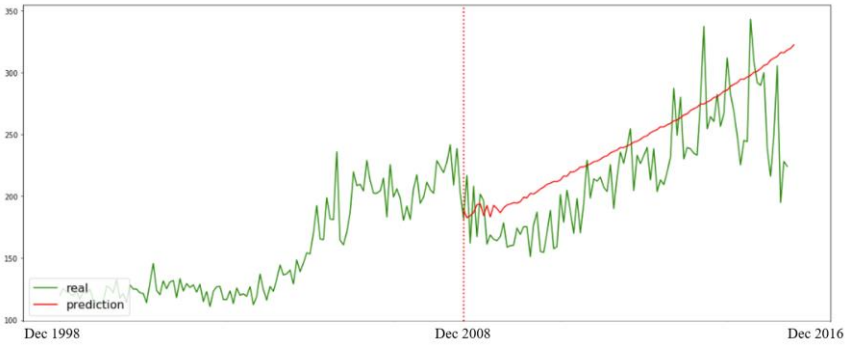


Figure 3-13. Eight years forecasting by LSTM

The final testing result is given in the figure. There are some common places for the two models. They are able to approximately catch the future trend of the data and could

not detect the decreasing trend very far near the year 2016. Compared to the seasonal ARIMA, LSTM did a better job on detecting the decreasing trend near the year 2008. This difference made the whole prediction of LSTM parallelly move downward, being closer to the actual level. The MAPE for seasonal ARIMA on the eight years prediction was 42%, while the LSTM was 24.6%. In term of the accuracy, the result is not promising but it led us to understand the behavior of the two prediction models for this long-term prediction. In reality, the model does not need to predict such a long period and the following two scenarios will be more practical.

3.4.2 Mid-term Prediction

The second scenario reduced the prediction time range and increased the frequency to update the model. For both time series and LSTM, four different models are trained respectively in sequence to predict the following two years indexes: trained on data from year 1998 to 2008 and predicted the indexes in 2008 to 2010, and then trained on data from year 1998 to 2010 and predicted the indexes in 2010 to 2012, etc. Another possible alternative was updating the start point of the training data as well. For example, in the second time period, instead of training on data from year 1998 to 2010, the models could be trained on data from year 2000 to 2010. It turned out this kind of procedure resulted in a worse out of sample prediction performance, probably because the dataset is not big and thus the loss of a piece of information undermined the model performance.

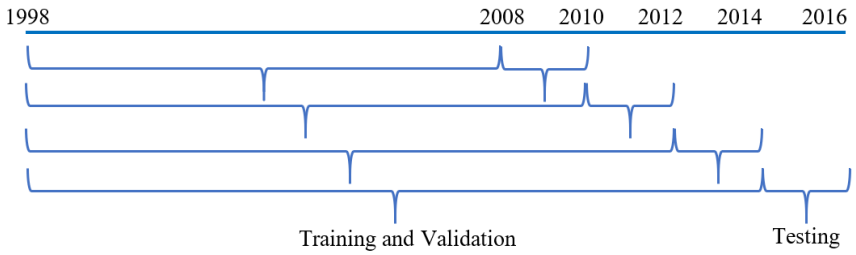


Figure 3-14. Split of the training and testing data

The training methods for two models are the same as discussed in the last section and thus are omitted here. For the time series model, it was found that adding the seasonal component in the midterm prediction helped to improve the out of sample accuracy, and the seasonality was set as 12 months. One explanation is that the data is volatile, so for the long-term prediction keeping the correct trend and making the average estimation would be the most conservative forecasting. In this experiment, for a short-term prediction, more volatile prediction increased the accuracy by catching the volatility, and the seasonality component added the extra volatility.

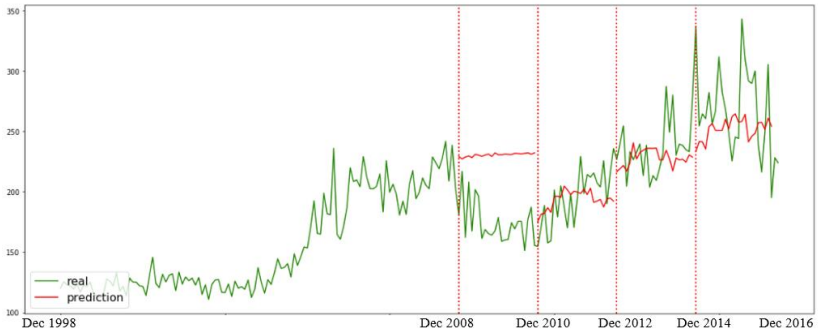


Figure 3-15. Two years rolling forecasting by seasonal ARIMA model

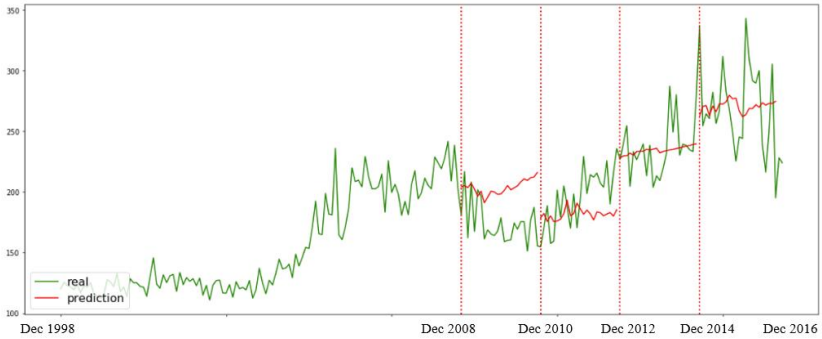


Figure 3-16. Two years rolling forecasting by LSTM

By comparing the two figures, the hardest task is still the prediction for year 2008 to 2010 because of the sudden trend change. LSTM showed a better ability to detect this change even though the error is relatively bigger compared to other prediction ranges. From the table, the accuracy of the LSTM is much higher than seasonal ARIMA for the prediction from year 2008 to 2010. The other three prediction periods are close and the time series model performed better than the LSTM in the second forecasting piece. Both models took the advantage of the new information to adjust the prediction for the following period.

Table 3-5. Out of sample prediction error for the midterm prediction (MAPE)

Prediction Year	08-10	10-12	12-14	14-16
MAPE of Time Series Model (%)	33.37	9.57	7.81	12.36
MAPE of LSTM (%)	18.51	10.27	6.38	11.39

3.4.3 Short-term Prediction

The short-term was defined as one-year, so there were eight seasonal ARIMA models and eight LSTM models. Each trained model only predicted for one years' indexes.

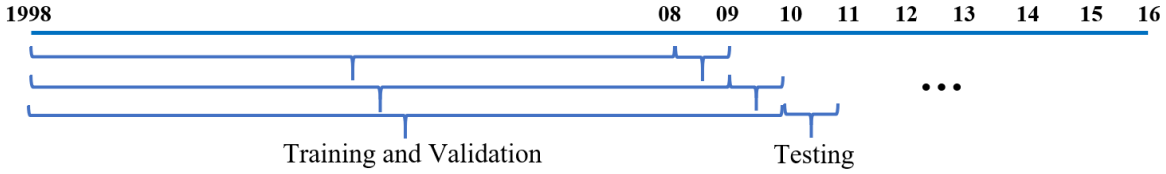


Figure 3-17. Split of the training and testing data

Two figures show the model performance visually. The seasonality in seasonal ARIMA was adjusted as 6 months, with a better validation and testing accuracy. LSTM demonstrated a better pattern recognition capability and in this one-year-prediction scenario, it looks more accurate than the seasonal ARIMA. Two tables support the conclusion from visual evaluation.

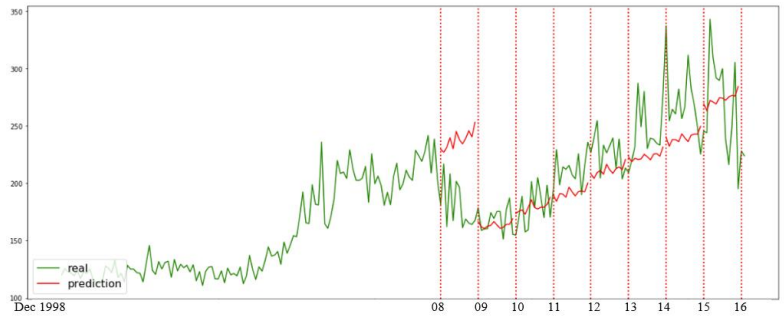


Figure 3-18. One year rolling forecasting by seasonal ARIMA model

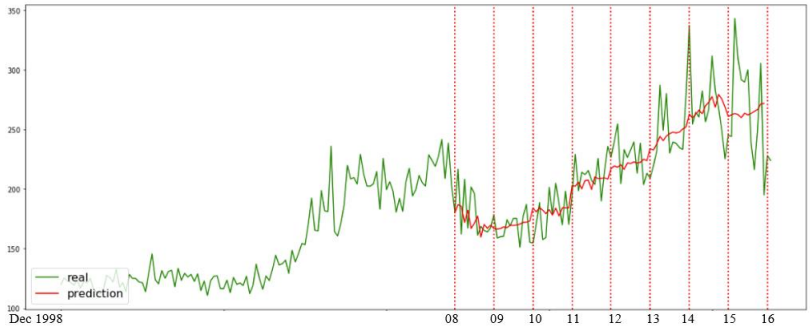


Figure 3-19. One year rolling forecasting by LSTM

Here two tables describe the model performance from different aspects. The table 3-6 is similar to the one in mid-term-prediction scenario and it calculates the error for each year's prediction, and therefore total eight metrices for time series and LSTM. As expected, most of the situation LSTM performed better than the seasonal ARIMA. Only one exception was for the prediction on year 2010 to 2011, where seasonal ARIMA model performed slightly better.

Table 3-6. Out of sample prediction error for the long-term prediction (MAPE)

Prediction Year	08-09	09-10	10-11	11-12	12-13	13-14	14-15	15-16
MAPE of Time Series Model (%)	33.59	5.73	7.80	9.51	8.29	8.83	12.43	14.99
MAPE of LSTM (%)	12.51	5.52	8.71	5.93	6.84	7.32	7.58	14.60

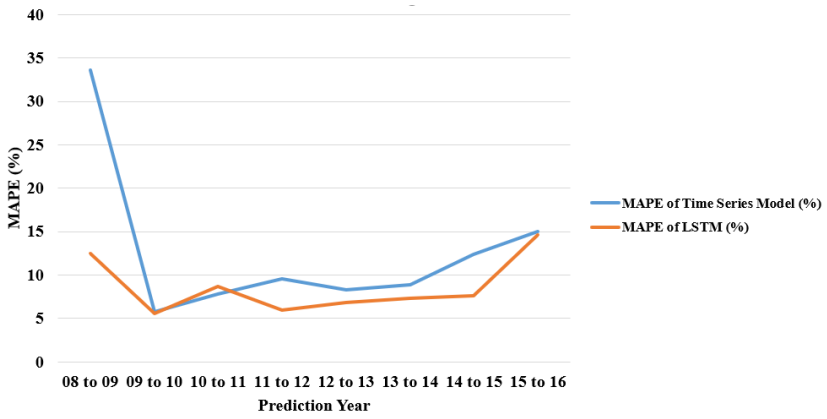


Figure 3-20. Comparison between LSTM and ARIMA

The figure 3-20 visually demonstrates the advantage of LSTM over the ARIMA. Form the performance of one-year prediction, LSTM is overall better than ARIMA, even though in some points LSTM is less accurate.

The table 3-7 evaluates the performance of each model on a specific future period. For example, the one-month MAPE for the time series model is the average of eight percentage error from the eight time series models, and each percentage error evaluates the deviation between the predicted next month index and the actual one. First, it seems that the time series method presents an evidently worse performance than the LSTM. This is because the first period prediction for time series model was overall bad (33.59 MAPE), and thus it increased the average level of every month's error. The conclusion we drew from this table is that the LSTM exhibits its memory sensitivity towards different time point. For example, in this dataset, it predicted very well in the first month, because it inferred that the nearby month would be closer to the current index. In contrary, the time series model attempted to make the prediction in an average level. LSTM works more like the human brain, which is differently sensitive to various time point.

Table 3-7. Out of sample prediction error for prediction a certain month ahead (MAPE)

Month Ahead	1	2	3	4	5	6	7	8	9	10	11	12
MAPE of Time Series Model (%)	16.48	8.76	9.24	8.55	14.43	12.96	11.54	10.68	9.42	8.73	15.46	14.32
MAPE of LSTM (%)	3.35	3.79	0.23	5.91	23.56	5.70	26.15	4.61	10.05	21.9	7.00	7.96

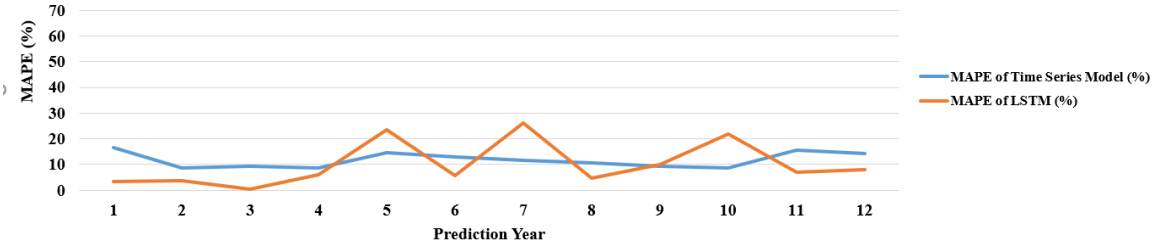


Figure 3-21. Comparison between LSTM and ARIMA

An auxiliary analysis was conducted to measure the effect of different training window. Even though the reason to choose the first breakpoint was introduced in the earlier section, a question was still left that if there would be any significant difference if the training window was changed. The original first training was from Dec 1998 to Dec 2008, and the following year's index was predicted. Here the end point of training window changes from the Aug 2008 to Nov 2008 month by month, and the following year was used as the test period. The results are displayed in the following table:

Table 3-8. Out of sample prediction error in different window end month

Training End Month	Jan 08	Feb 08	Mar 08	Apr 08	May 08	Jun 08	Jul 08	Aug 08
MAPE of Time Series Model (%)	23.51	21.46	22.74	20.69	25.15	29.65	39.64	41.37
MAPE of LSTM (%)	31.79	27.46	29.15	25.94	28.31	29.45	41.15	40.16
Training End Month	Sep 08	Oct 08	Nov 08	Dec 08	Jan 09	Feb 09	Mar 09	Apr 09
MAPE of Time Series Model (%)	39.74	36.92	35.37	33.59	24.92	25.26	21.46	18.42
MAPE of LSTM (%)	41.93	39.48	25.72	12.51	12.87	11.92	11.88	9.53

Table 3-8 continued

Training End Month	May 09	Jun 09	Jul 09	Aug 09	Sep 09	Oct 09	Nov 09	Dec 09
MAPE of Time Series Model (%)	12.25	8.52	7.41	8.74	7.15	7.85	8.14	8.23
MAPE of LSTM (%)	8.41	8.12	6.23	7.25	7.24	6.27	5.81	4.94

From the table, the ARIMA exhibits unsensitivity to the possible decrease trend and until the end point was cut at January 2009, the ARIMA starts to modify the prediction to accommodate the change. LSTM detects the change earlier than ARIMA because it starts to improve the model prediction when the end point was cut at November 2008. This analysis supports the rationale of selection the December 2008 as the first cut point.

The last experiment was conducted to see the performance of a naïve prediction model, linear regression. The difference between linear regression and ARIMA is that ARIMA considers both the autocorrelation and partial correlation, while the linear regression only considers the autocorrelation. In order to make the linear regression model comparable to the ARIMA and LSTM, it is trained using the data that maps 24 former months to the 1 following month, as linear regression could not make multiple predictions at one time. However, the same model proposed in LSTM applied here in linear regression: in out of sample prediction, the input will be updated using the previous prediction. The results are displayed in the following figure and tables.

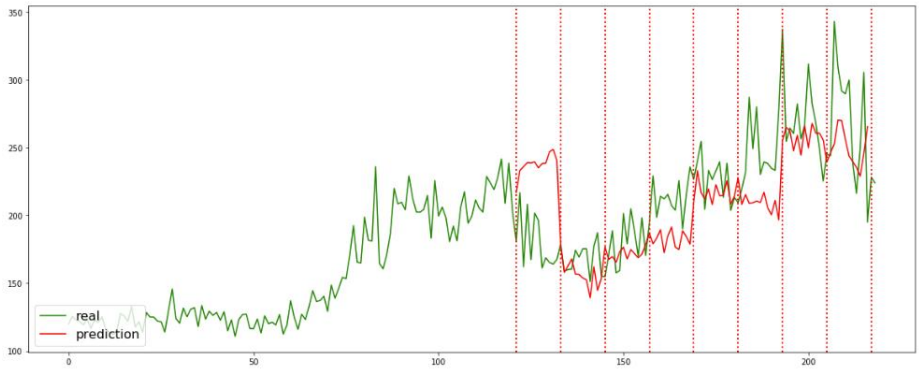


Figure 3-22. One year rolling forecasting of linear regression

Table 3-9. Comparison between three methods

Prediction Year	08-09	09-10	10-11	11-12	12-13	13-14	14-15	15-16
MAPE of Time Series Model (%)	33.59	5.73	7.80	9.51	8.29	8.83	12.43	14.99
MAPE of LSTM (%)	12.51	5.52	8.71	5.93	6.84	7.32	7.58	14.60
MAPE of Linear Regression (%)	33.54	7.55	8.38	13.42	6.56	14.54	8.87	12.79

From the results we can see that the linear regression is the worst method in terms of the out of sample prediction accuracy, but in 15-16 it is the most accurate. This was partially explained by the fact the linear regression is not stable and the prediction is most volatile, so by accident the big volatility matches the actual data.

After the experiments in three scenarios, a comprehensive comparison was conducted between a time series model and the LSTM. From the perspective of the prediction accuracy, LSTM is generally better than seasonal ARIMA, while in several cases ARIMA performed a little better. LSTM is better at detecting the trend change. In terms of the training procedure, the LSTM is more complex, but the advantage is it has no requirement on the raw data, while time series model needs a stationary input. For the training time, the LSTM is longer than that of time series model, but it is still acceptable (around 10 seconds), so this will not be problematic for LSTM to be applied in practice.

3.5 Conclusions

Except the further need for exploring a method that could make a more accurate prediction, this research discussed two important considerations when assessing a prediction model. The first consideration is that by initial analysis of the raw data, researcher should have a brief estimation of the prediction difficulty and thus set a suitable expectation for the prediction models. Evaluating a model only based on the error metrices might result in an overestimation. The second consideration is that when evaluating the time effectiveness of a prediction model, extra attention needs to be paid to check if there is any future information used in the test process. If that is the case, the time effectiveness of the model might be overestimated. This research applied the sequence to sequence (seq2seq) model based on LSTM units in predicting the HCCI. The seasonal ARIMA was the benchmark for the comparison. Three different prediction scenarios, including the short-term prediction (1 year), mid-term prediction (2 years), and the long-term prediction (8 years) were simulated to demonstrate the different performance of the time series model and LSTM. The research not only focused on the model prediction power, but also laid the emphasis on the potential industry applicability by discussing the training time.

4. TREND AND VARIATION ANALYSIS OF THE UNIT PRICE BIDS OF RESURFACING PROJECTS IN GEORGIA

4.1 Introduction

This part of research extends the exploration from univariate prediction to the multivariate version. In practice, it is necessary to find out the explanatory variables which could provide extra information to better predict the price or cost. A research project was supported by Georgia DOT related to investigating the variables to predict the unit price bids of resurfacing projects. A powerful non-linear feature algorithm, namely Boruta analysis was explored to find out the useful features and it could be visually displayed using the partial dependence plots. An ensemble learning model was developed using the selected features to accurately predict the unit price bids.

4.2 Research Objective

There are three major research objectives:

First, considering the significant variability of the unit price bid, there is a need to develop a robust prediction method that is capable of providing reasonable forecasts under different conditions. The main goal is to enhance the accuracy of prediction under different conditions.

Second, there is also a need for developing a prediction model that can utilize and detect the information embedded in other variables to enhance the prediction of highway

construction cost. It is important that the prediction method has the capability to handle a wide range of features that have potentials to improve cost estimation. There is a need for an efficient forecasting algorithm to select the best subset of features that provide the desired level of accuracy in predicting the cost.

Third, a desirable forecasting model needs to: (a) Handle a large number of variables efficiently and effectively; (b) Work with both numerical and categorical variables; and (c) Deal with missing data points that is unfortunately a common problem in highway construction cost analysis.

4.3 Research Methodology

4.3.1 Research Framework

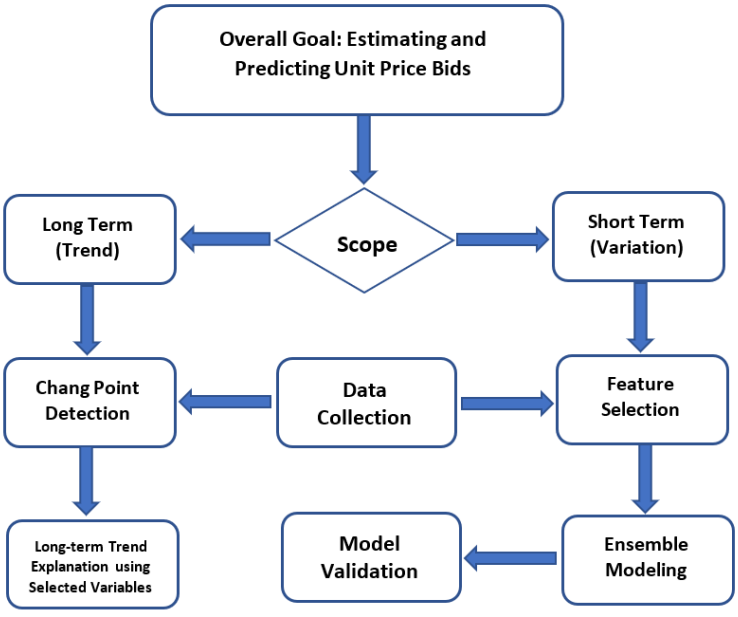


Figure 4-1. Research formework for the part two

The key portion in this research is the data collection. The quality of the data to some extent determines the upper limit of the prediction models. To roughly analyze the

long-term trend of the cost, a non-parametric analysis framework will be used for detecting the change point and estimate the trend. To model the variation, more complex features are required, and the feature selection algorithm helps to choose the optimal set of variables. The ensemble learning model is supposed to produce the accurate forecasting and the model validation will show the comparison between proposed model and the existed models.

4.3.2 Data Collection

This research devoted a significant part of effort in data collection by The Georgia Tech ESBE lab. Several different data sources were combined in the collection process. Collected data could be divided into project-specific features and the macro-economic level indexes. The relation between the project specific attributes and the bid prices was researched in many studies such as the one from Shane et al., who argued the significant relation between duration, complexity, and the number of bidders, on the construction cost escalation (Shane et al. 2009). The macroeconomic indicators could also explain the variation of the construction cost. Akintoye et al. detected the leading indicator of construction cost and they found that the unemployment level and the industrial production could be used to explain the variation of construction cost (Akintoye et al. 1998). Ng and his colleagues built the model to predict the change direction (either up or down) of Hong Kong tender price index with eight local indicators, such as gross domestic product (GDP) and unemployment rate (Thomas et al. 2000). Wang and Liu focused on the average bidding price of resurfacing projects in Kentucky to see the relation of bidding price to four factors, including local asphalt price index (Wang and Liu 2012).

The major data source was the BidTabs Oman system, the BidExpress online system, and the GDOT. The Oman system is a widely applied software used by many DOTs to store the let bid data. The available data includes but is not limited to: Job Number, Pay Item, Quantity, Unit, Unit Price, Extension Amount, Bid Date, County, Bidders, District, Position, Low Bidder and the Total Contract Price. The public dataset is available from the following websites: GDOT Bid Express, U.S. Energy Information Administration, and Bureau of Labor Statistics.

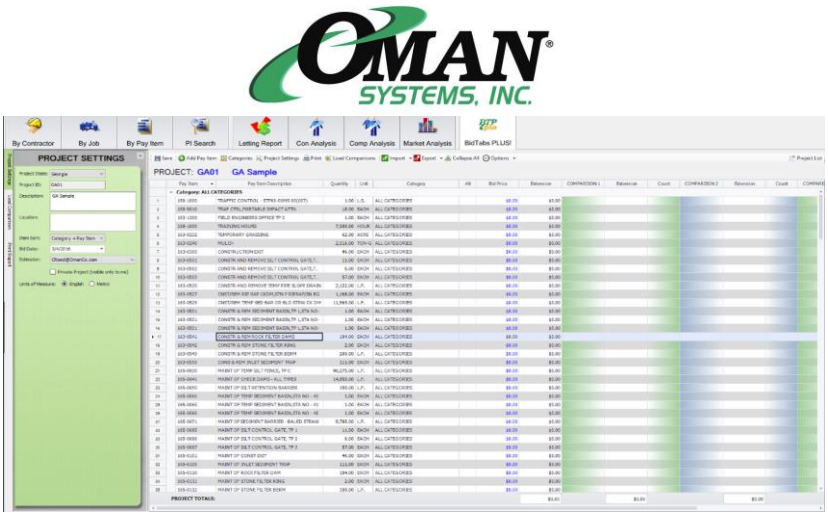


Figure 4-2. The interfance of Oman System

(Figure from: http://omanco.com/index.php/product/bidtabs/)

Another useful database is the Bid Express's Tabulation of Bids, which is currently used by GDOT to report the bid submission and letting information. The available data includes but is not limited to: Contract ID, Letting Date, County, District, County, and Area Office, Project ID, Line No, Item ID, Item Description, Quantity and Units, Bidders, Unit Price submitted by all bidders, Extension Amount, Total Contract Price submitted by all

bidders, Position (Rank), Low Bidder, Contract Time (MM/DD/YY, Completion Time), and the Contract Description.

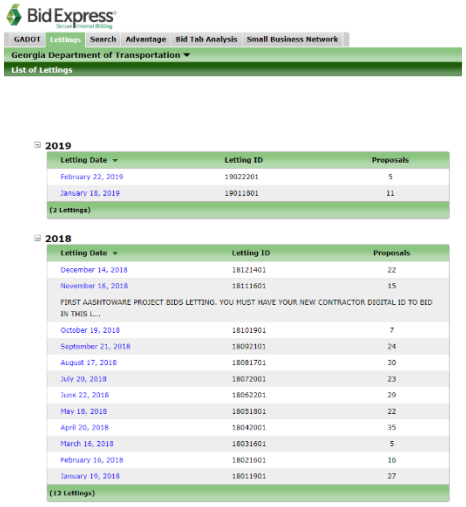


Figure 4-3. The screenshot of Bid Express online system

The ESBE lab, who worked with the Georgia Department of Transportation, finally collected the average unit price bids data and the data ranges from January 2008 to January 2016. Data cover over 1400 resurfacing and widening projects' lowest tender price and have been normalized to unit price (dollars per ton). Among these projects, very few of them are design build projects and most of them are design bid build projects. All projects use fixed-price contracts and the following quantity was proposed by each contractor. Even though each project has several submitted bids, only the winning bids were considered. In Georgia, the most common asphalt line items for resurfacing and widening are hot mix recycled concrete (i.e., 9.5 mm, 12.5 mm, and 19 mm Superpave), a mix of reclaimed asphalt pavement, reclaimed asphalt shingles, virgin aggregate, hydrated lime and neat asphalt cement (Baek 2018, Floy et al. 2013).

Submitted unit price bids are subject to substantial variations over time and from project to project, which make prediction hard. It can be seen from the Figure 1 that for example in June 2013 the difference between the greatest and the lowest submitted bid prices can be as large as $106.17/ton, i.e., the greatest bid price is about 3 times larger than the lowest bid price.

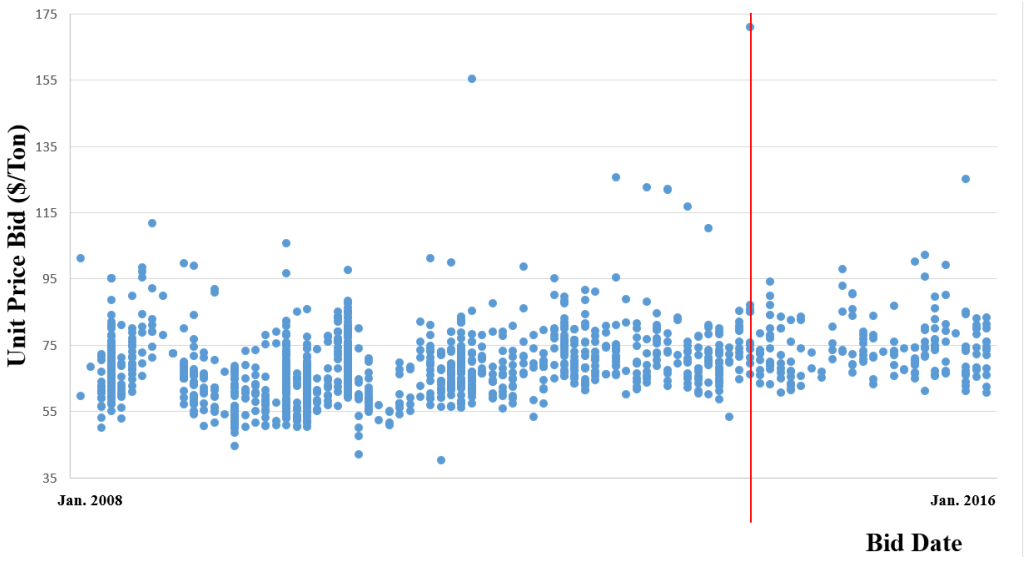


Figure 4-4. Submitted unit price bids for resurafacing projects

There is also a considerable variation in the monthly rate of change of the average value of submitted bid price. For example, the average unit price bid went down by over 30% from 84 $/ton in October 2008, to 57 $/ton in May 2009. (Fig. 2).

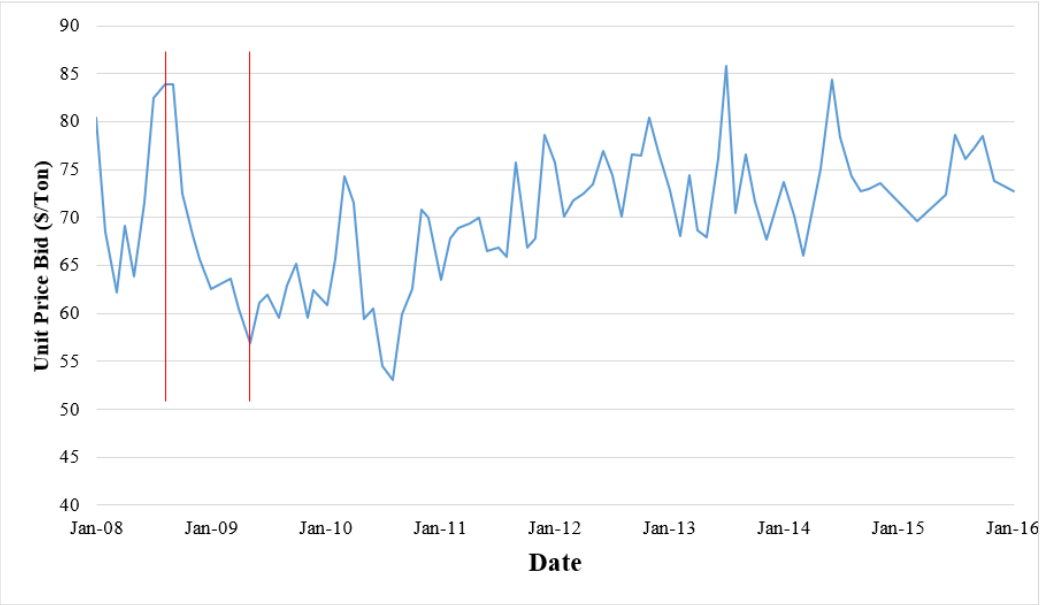


Figure 4-5. The monthly average of submitted unit price bids

For the purpose of implementing the ensemble machine learning model and making an accurate prediction, more features are required to forecast the unit price bids. The Georgia Tech ESBE lab performed the data collection to get the project specific attributes and macroeconomic indexes.

All 57 features could be categorized into the following groups:

1. Features describing project characteristics: Project duration, quantity of the bid item, total bid price, project length, number of pay items, and the number of bidders.

This group of attributes describe the basic information of a project. The project duration is measured by the interval between the notice to proceed (NTP) and completion date. The quantity of the bid item is the volume of asphalt line item in the submitted bid. The total bid price is the lowest submitted bid amount. Project length is the total pavement length of the project. The number of pay items tells us the different work type in a project. It is around 30 to 40 for each project.

2. Features related to project location and its distance to major supply sources for critical materials: Terrain of the project, district of the project, number of asphalt plants within 80.5 kilometers, hauling distance between asphalt plant and project location, and hauling distance between quarry and asphalt plants.

In the later research, this part was turned out to be with top explanatory power to predict the unit price bids. There are seven regions in Georgia as shown in the figure. Each region gathers the closed counties to form a relative independent area, for the convenience of government management and regulation.

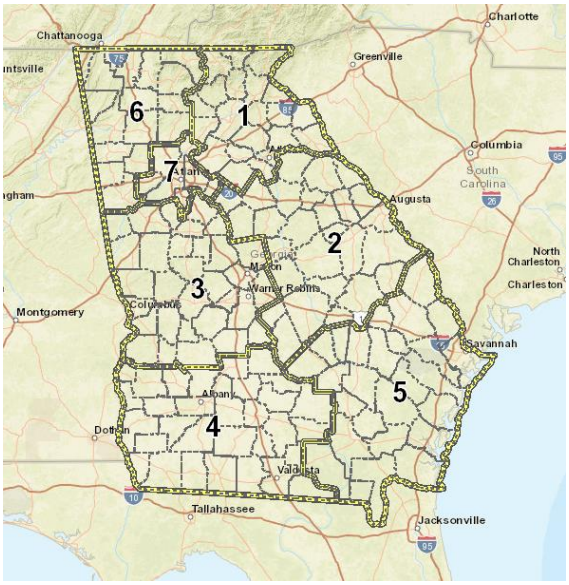


Figure 4-6. Screenshot of region map in Georgia, figure from GeoPi

Geographically, Georgia is divided into four areas: coastal, flat, mountainous and the rolling. There is no former research who found the direct relation between the terrain type and the unit price bids, but based on the literature review of authors, the impact could stem from following parts: different terrain is related to the accessibility of the resource and materials; the temperature is different in various terrains and it impact the difficulty of the construction process (Baek 2018).

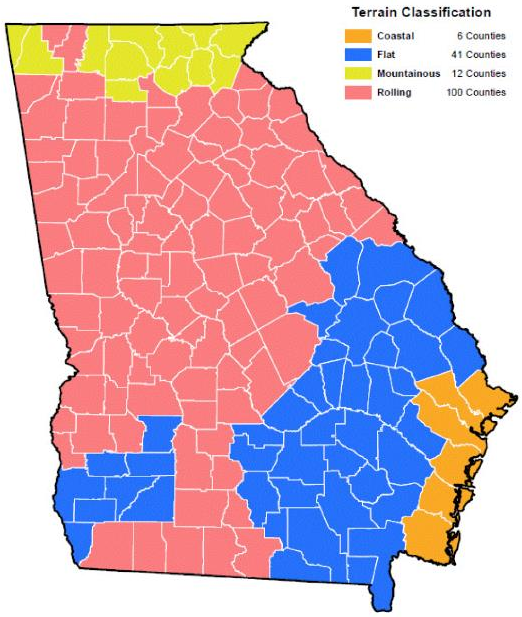


Figure 4-7. Terrain map of Georgia

(Figure from "Best Practices for Budget-based Design", Ashuri, Baek and Li)

As mentioned to the resource and material accessibility in the terrain discussion, the number of asphalt plants within 80.5 kilometers exactly depicts the difficulty to get the main material of highway construction projects. A research report states that the appropriate distance between the construction site and the plant should be within 80.5 kilometers (Baek 2018).

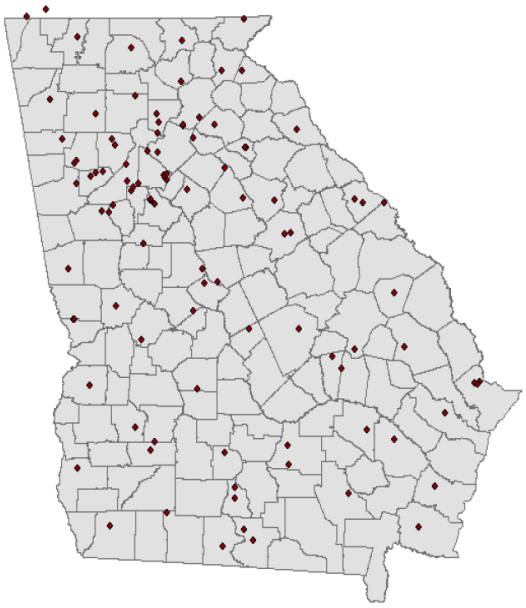


Figure 4-8. Asphlat plant map of Georgia

(Figure from "Best Practices for Budget-based Design", Ashuri, Baek and Li)

3. Features representing level of activities in the local highway construction market: Total monthly asphalt volume of resurfacing and widening projects awarded in the same month at the level of the county, total number of resurfacing and widening projects awarded in the same month at the level of the county, total dollar value of awarded projects at county level, total number of projects awarded in the same month at the state level, total dollar value of projects awarded in the same month at the state level, and the total asphalt volume of projects awarded in the same month at the state level.

The meaning of each feature is intuitive. This group describes the level of activities in either state or county level. It is assumed that the area with the more intensive projects would be equipped with a more convenient infrastructure and material source, and thus might exhibit a lower unit price bids. All features in this group was collected from the BidTabs database.

4. Features representing construction market conditions: Architecture Billings Index (ABI), Building Permits for New Residential Construction, Building Cost Index, Common Labor Index, Construction Cost Index, Equipment Operator Wages (Paving), Fails Management Institute (FMI) Nonresidential Construction Index, Asphalt Cement Price Index, Gross Domestic Product of the Georgia Construction Industry, Housing Market Index, Labor Productivity, Material Price Index, National Highway Construction Cost Index, Number of Establishments in Private Construction Industry, Number of Hires, Producer Price Index (Construction machinery manufacturing), Producer Price Index (Construction sand and gravel mining), Skilled Labor Index, Turner Construction Cost Index, Value of Construction Put in Place (Pavement), Value of Construction Put in Place (All construction), 12-Month Percent Change of Asphalt Cement Price Index, 12-Month Percent Change of Gross Domestic Product of the Georgia Construction Industry, 12-Month Percent Change of the Number of Hires, and 12-Month Percent Change of Value of Construction Put in Place (Pavement).

These are the related economic indexes that retrieved from several data sources including the U.S. Energy Information Administration, U.S. Bureau of Labor Statistics, U.S. Census Bureau, etc.

5. Features representing overall macroeconomic conditions: Average weekly wage (all industry), Consumer Price Index (South), Dow Jones Industrial Average, Inflation rate, Population, Producer Price Index (Gasoline products), Producer Price Index (Steel mill products), Producer Price Index (No. 2 diesel fuel products), Producer Price Index (Crude petroleum products), Unemployment, and the 12-Month Percent Change of Unemployment.

As the highway construction projects are labor intensive, all the above attributes were collected with the assumption about the relation. Some of the attributes have been applied in the former research as the leading indicators to predict the NHCCI, such as the average weekly wage (Shahandashti and Ashuri 2015).

6. Features representing oil market conditions: Crude Oil Price of West Texas Intermediate (WTI), Diesel Retail Price, Fuel Price Index, 12-Month Percent Change of WTI, and the 12-Month Percent Change of Fuel Price Index.

Crude oil and fuel are supposed to be two of the most significant raw materials in highway construction cost index. The crude oil price was collected form the Federal Reserve Bank database and it is the spot price of unrefined petroleum product measured in dollars per barrel. Diesel retail price is the local value in Georgia, which could reflect more about the local market.

The majority of the above features are time related, and others are project specific. To overcome the problem related to "cheating prediction", for the time-related features, the one-month lag is created between the predictors and the response variable (unit price bids) for the purpose of training a predictive model. For example, the Fuel Price Index (a time-related feature) in this month will be used to predict the unit price bids in the next month, and this is called roll forward cross validation in machine learning. Not all of the above features will be used to train the machine learning model, and that detailed discussion is in feature selection part.

4.3.3 Non-parametric Framework for Trend Analysis

The procedure of this part is first detecting the change point of the price time series data. The meaning for each separated sequence is that the trend is stable based on statistics. The test of trend is calculated using the Man-Kendall test and the estimation is computed with the Theil-Sen estimator. Second, for each stable period we analyze the trend of other explanatory variables in same ways and find out the trend relation between explanatory variables and the unit price bids. All the statistic methods applied are the non-parametric methodologies, and thus under no limitation to the raw data. The analysis framework is supposed to be applicable to any similar problem.

In this research scenario, the change point detection is significant to get the stationary sequences which could be further analyzed to find the trend. In the above figure, six types of change point are displayed. Type c and type e are two of most frequent change point in dataset of this research. Change point c means the mean (median) shift with the decrease of volatility; change point e means the change of slope level on long term linear trend.

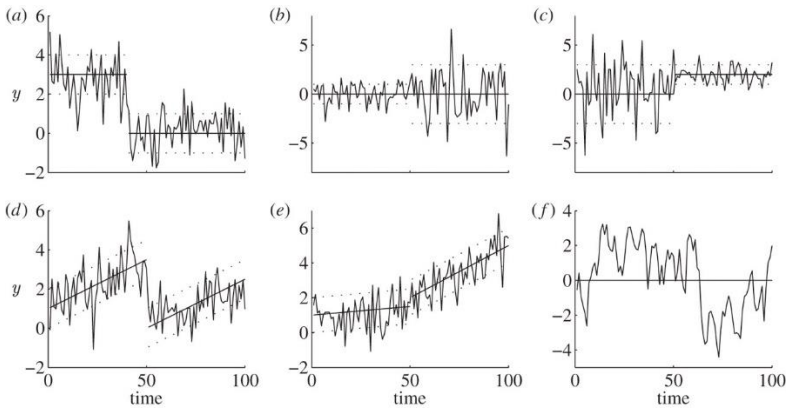


Figure 4-9. The function of change point detection

In order to realize the task, this research referred to the work of James and his colleagues published in 2014. They proposed an innovative breakouts technique which employs Energy Statistics to detect breakouts. The technique uses robust statistical metrics, viz., median, and estimates the statistical significance of a breakout through a permutation test. As mentioned by the authors, this is the first work which addresses breakout detection in the presence of anomalies (James et al. 2014).

Suppose that we are given the following time series, Z1, Z2…Zn consisting of independent observations. A breakout is characterized by a value $\gamma \in (0, 1)$ such that observations {Z1, Z2... Zγn} have distribution function F, and observations {Zγn+1, Zγn+2 … Zn} have distribution function G. Furthermore, it is assumed that $F \neq G$. In order to determine if the observations in the provided time series are identically distributed, we perform the following hypothesis test:

H0: $\gamma = 1$

HA: $0 < \gamma < 1$

If the null hypothesis of no breakout is rejected, we must then also return an estimate for the breakout location (James et al. 2014).

After we get the stationary data pieces from the above change point detection algorithm, statistical test is required to see if there is a monotonic linear trend and fit a suitable line to the trend data to quantitatively depict the trend.

The purpose of the Mann-Kendall (MK) test applied in this research (Kendall 1975) is to statistically assess if there is a monotonic upward or downward trend of the variable of interest over time. A monotonic upward (downward) trend means that the variable consistently increases (decreases) through time, but the trend may or may not be linear. The MK test can be used in place of a parametric linear regression analysis, which can be used to test if the slope of the estimated linear regression line is different from zero. The regression analysis requires that the residuals from the fitted regression line be normally distributed; an assumption not required by the MK test, that is, the MK test is a non-parametric (distribution-free) test.

The trend estimation is calculated through the Theil-Sen estimator, which is a method for robustly fitting a line to a set of points (simple linear regression) that chooses the median slope among all lines through pairs of two-dimensional sample points (Theil 1950). This estimator can be computed efficiently and is insensitive to outliers. It can be significantly more accurate than non-robust simple linear regression for skewed and heteroskedastic data, and competes well against non-robust least squares even for normally distributed data in terms of statistical power (Wilcox 2012). It has been

called "the most popular nonparametric technique for estimating a linear trend" (El-Shaarawi and Piegorsch 2001).

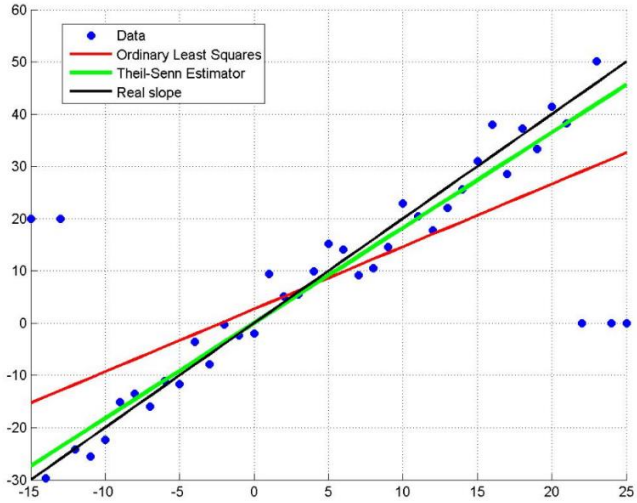


Figure 4-10. Theil-Sen estimator
(Figure from https://www.mathworks.com/matlabcentral/fileexchange/34308-theil-sen-estimator)

As demonstrated in the figure, for a noisy data set that generated from a central trend in black line, the Theil-Sen estimator fits a line close to real line while the ordinary least square regression deviates a lot due to some outliers. In this research, bidding price in many days is like the outliers in this example and thus the Theil-Sen estimator will give a more robust line fitting.

4.3.4 Feature Selection

A contribution of this thesis is proposing a model that could deal with missing values. From the literature review, four methods are frequently used to process missing value: delete rows; replacing missing values with mean or median and predicting the missing values. Considering that each row is a project, so the information is significant,

then the first method is not a good selection. Replacing the missing value with mean or median and the third method has the similar result, but the third model is more convincing, because it uses all the information from the dataset to predict the missing values, instead of only considering the missing columns. This research uses the third method even though in this research, there is no difference in terms of in sample accuracy. Two reasons result in the phenomenon: first, there are only 5 missing cases among 1400 observations and the attributes with missing values are not significant based on the following analysis. However, the choice of the methodology us significant for the stable application when there is a larger number of missing values.

Feature selection is required for two reasons: first, too many features will slow down the speed of machine learning algorithms; second, many algorithms demonstrate a decrease of accuracy when the number of variables is higher than the optimal one (Kohavi and John 1997). Both the feature selection and model training/testing is conducted in R studio. The Boruta feature selection algorithm is a unique aspect of the proposed ensemble learning method that facilitates the selection of the most relevant features with the greatest advantage for enhancing the cost prediction. The Boruta feature selection algorithm is particularly useful in the context of forecasting highway construction cost as it can select the most important features form a large number of variables to improve cost forecasting. Especially, the selection is challenging as the relationships among the features and the unit price bid are complex and nonlinear. As shown in the table, where the features are ranked based on their Pearson correlation, the largest correlation between the unit price bid and any of the variables in the dataset is 0.44 (the correlation between quantity and the unit price bid), while the rest of correlation coefficients are less than 0.41. Only half of the

features have the correlation bigger than 0.2. Therefore, the application of the conventional multiple linear regression method is limited in the context of the dataset under study in this research. The Boruta algorithm is shown to be a successful method to find a best subset of the features to develop a forecasting model with an outstanding performance as far as accuracy.

Table 4-1. Pearson correlation analysis results (partial)

DJIA	0.44
Consumer Price Index (south)	0.41
Architecture Billings Index	0.40
GA Asphalt Cement Price Index	0.39
GDP (levels) (12-month percentage change)	0.39
HMI (South)	0.37
FMI NRCI	0.37
Building Permits for New Residential Construction	0.37
Year	0.36
ENR Material Price Index	0.36
GA Fuel Price Index	0.34
ENR Building Cost Index	0.34
Diesel Retail Prices, Lower Atlantic	0.33
PPI, Construction machinery mfg	0.32
ENR Construction Cost Index	0.32
ENR Skilled Labor Index	0.32
PPI, Gasoline, Fuels and related products and power	0.31
ENR Common Labor Index	0.31
PPI (Construction sand and gravel mining, National)	0.30
Equipment Operator Wages, Paving, Mean hourly wage, Georgia	0.30
PPI, Steel mill products, Metals and metal products	0.30
PPI, No. 2 diesel fuel, Fuels and related products and power	0.29
PPI, Crude petroleum (domestic production)	0.25
Flat	0.24
Coastal	0.23
Inflation rate	0.23
Crude Oil Price	0.22
Turner Cost Index	0.20
Total Bid Price	0.20

The way to determine if an attribute is unimportant is based on the two-tailed hypothesis test within the algorithm. The detailed procedures are developed as follows:

1) A copy is made of each explanatory variable. The values in these copies are permuted to remove any correlation with the target variable. The copies are called shadow variables.

2) A random forest model is fitted on this expanded data set.

3) For each variable (the original and the shadow), the Z-score of the loss in accuracy is calculated. The Z-score is the average loss divided by the standard deviation.

4) Record the z-scores of the original attributes that are higher than the maximum z-score of the shadow attributes.

5) Repeat the above steps numerous times. Original attributes that are significantly--statistically--higher than (the hypothesis tests are finished within the algorithm with a level of significance of 0.01, for the purpose of testing if a z score from one attribute is significantly different than the maximum one) the maximum z-score of shadow attributes are deemed relevant to the prediction. Attributes that are significantly below the maximum z-score of shadow attributes are deemed not relevant.

All of the above procedures are finished within the algorithm and result of hypothesis test is reflected in z-score, or “importance measure”.

Another important technique related to the feature selection is partial dependence plot. This tool is totally based on the result of Boruta analysis but it exhibits the result in another way. First proposed by Friedman in 1999, partial dependence plot is a powerful tool to visually show the non-linear relation between any independent variable and the

dependent variable (Friedman 2001). A complex machine learning algorithm is always compared to the linear regression, which is a comparison between the non-linear model and the linear model. One of the biggest advantages of linear regression is the explanatory power of the model. The coefficient of one feature could be easily explained as by changing one unit of the feature, a corresponding coefficient amount of dependent variable will be changed. For the non-linear model, such explanation is lost due to the more complex non-linear relation. Friedman first proposed this concept in the gradient boosting algorithm, which is a tree based boosting algorithm. His idea on the partial dependence plot is described as follows: consider we train an arbitrary machine learning model with a given dataset. This dataset includes N observations of a response variable, along with p features, the model generates predictions of the form:

$$\hat{y}_k = F(x_{1,k}, x_{2,k}, \ldots, x_{p,k})$$

here the F(.) is the fitted model. Friedman defined the partial dependence of one feature j as:

$$\varphi_j(x) = \frac{1}{N}\sum_{k=1}^{N} F(x_{1,k}, \ldots, x_{j-1,k}, x, x_{j+1,k}, \ldots, x_{p,k})$$

The mathematical form is simple, but this creative idea endowed the explanatory capability to machine learning algorithms (CRAN. 2019). It works like the marginal distribution. For one value of an attribute, the average of all the function values denotes the marginal value for this feature, given the value. The strength of this method could also be reflected when compared to the Pearson correlation analysis, which is the quantification of the marginal effect in linear models. In linear regression, the dependent variable is

analyzed directly to each independent variable and then decide if there is any linear correlation. The biggest shortage of this scenario is that when there is significant linear correlation, the relation might caused by other variables; in contrary, when there is no significant linear correlation, that might due to the noise caused by other independent variables. The "contribution" of each independent variable to the dependent variable needs to be considered after moving out impact of others.

The algorithm was proposed in the context of tree boosting, but it could also be applied in other machine learning algorithms such as support vector machines. In the current programming environment, partial dependence plot is provided only in tree-based algorithms such as random forest and gradient boosting, but it is intuitive to be implemented based on the formula.

4.3.5 Ensemble Learning

Machine learning algorithms are used to predict the bidding price in this research. Compared to related work, this research proposed an ensemble learning model, which comprises four machine learning algorithms, to provide a more accurate bidding price prediction.

Dietterich defined ensemble methods as the learning algorithms that construct a set of classifiers and then classify new data points by taking a weighted average vote of their predictions (Dietterich 2000). Two key points are mentioned in this definition: (1)

ensemble methods are composed of more than one machine learning algorithm and (2) the final results are the weighted average of each prediction from different algorithms.

The technic of ensemble learning has been applied in the construction industry for several years, even though the number of cases is limited compared to the application of single machine learning algorithms. Williams and Gong applied the text mining and ensemble learning classifier to predict the cost overrun of the project based on the contract document. They found that the stacking model (ensemble learning model), which is composed of Ridor, K-Star, and RBF neural network, produced the best result, even though the accuracy was about 44% (Williams and Gong 2014). Chou and Lin focused on predicting if there will be disputes in public–private partnership projects based on some preliminary projects parameters. They found that the ensemble learning technique, which was composed of a support vector machines, artificial neural network, and the decision tree, provided the best prediction result, which was about 84% (Chou and Lin 2012).

In this research, the innovation is that the ensemble model is composed of two layers of prediction models, as shown in Figure 3.

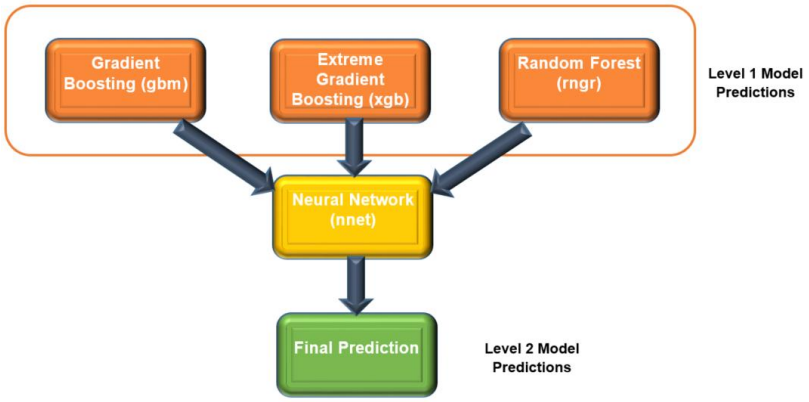


Figure 4-11. The model structure of the ensemble learning

Three machine learning algorithms are developed in the first-level model training: gradient boosting (gbm), extreme gradient boosting (xgb) and random forest (rngr). Each of these three models is capable of both doing classification and regression.

The reason to choose gbm and xgb is because they have capabilities from both gradient descent and boosting. Some other machine learning algorithms were compared with the performance of these two algorithms and they stood out for the following three reasons: first they are capable of dealing with both numerical and categorical attributes; second, there is embedded parallelization manipulation in the gbm package in R that makes the computation faster than other algorithms when dealing with the same dataset; third, they can deal with sparse matrix which means they are stable even when facing with missing values. Gradient descent is an efficient and well-formed optimization algorithm aimed at finding the optimal solution of loss function through simultaneously updating the value in each direction of gradient, and boosting helps to improve the accuracy of gradient descent by giving the poor case (high error) with higher weights and renewing the model. In each stage of gradient boosting, it introduces a weak learner to compensate for the shortcomings of the existing weak learners. The main difference between gradient boosting

and extreme gradient boosting is that the latter uses a more regularized model formalization to control over-fitting, which gives it better performance. The comparison result proved this point later.

The main reason to choose random forest (rngr) is because the tree model is to cooperate with the Boruta analysis to further conduct the feature selection. Secondly, it is one of the most accurate machine learning algorithms in practice; in addition, it can handle a lot of input variables without variable deletion. In this research, all the chosen features could provide information, more or less, to predict the bidding price. Random forest will try to use all of them to perform the forecasting.

In most built-in machine learning algorithms in R, the accuracy metric for model training is root mean square error (RMSE), which is calculated by:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} {e_i}^2}{n}} \quad 1.$$

Compared to the mean absolute error (MAE), RMSE amplifies and severely punishes large errors. The RMSE reported from the calculation for gradient boosting, extreme gradient boosting, and random forest are 0.12, 0.09, and 0.08, respectively. These results only mean that the trained models are acceptable in the perspective of good fit to the training data. No conclusion could be drawn here concerning how good these models are to predict the future bidding price. The three models exhibit different prediction characteristics: gradient boosting leads to the result with a higher variation compared to the other two algorithms, but it better captures the changing trend of the bidding price.

In the second level of ensemble modeling, a neural network is selected to produce the final predictions. Compared to linear regression, the neural network is good at modeling complex nonlinear relation and is more widely applied. Three components of a neural network are input layer, hidden layers, and output layer. In this research, the input layer is composed of three nodes, which are the corresponding result calculated from gradient boosting, extreme gradient boosting, and random forest. The output layer has only one node: the prediction of the value of unit price bids. The difficulty is to determine the structure of the hidden layer(s): how many layers are there in the neural network and how many hidden units are there in each layer. The number of hidden layers is decided to be one because there are only three input nodes, making the simplicity reasonable. The best number of hidden nodes (units) is determined through cross-validation by attempting a different number of units. In this research, six attempts are made to train the hidden layer with 1, 3, 5, 7, 9, and 11 units. Through the iteration, 9 turned out to be the optimal number, which makes the neural network get the smallest RMSE.

. The structure of the neural network is displayed as follow:

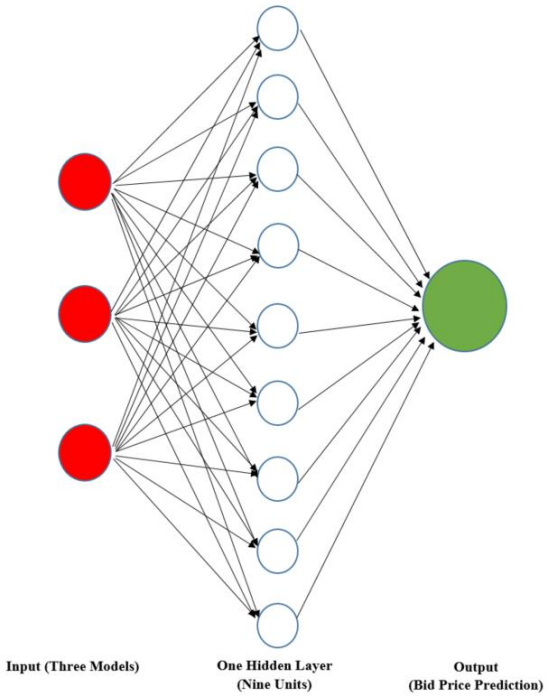


Figure 4-12. Selected structure for neural network

A simplification of the model is that the three input components of the neural network are equally weighted and it was proved to be able to produce a satisfactory prediction accuracy. There are some other more complicated and advanced weight decision methods that could be tested in future research, such as Bayesian technic, boosting, etc.

4.4 Results and Discussion

4.4.1 Trend Change Analysis based on the Non-parametric Framework

The average unit bidding price data is collected by Georgia Tech ESBE lab, ranging from Jan. 2008 to Jan. 2016. Data covers about 1400 resurfacing and widening projects' lowest tender price and has been normalized to unit price dollar per ton.

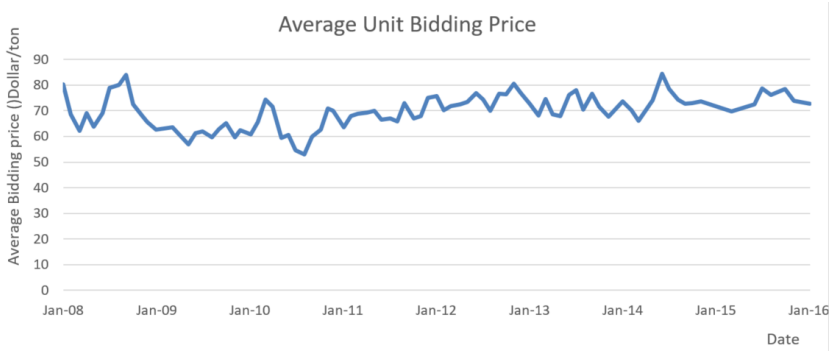


Figure 4-13. Original Dataset

Based on the comprehensive literature review, the following six factors are chosen as the relatively closely related indicators of bidding price.

Table 4-2. Six Selected Indicators

Indicators	Full Description
Georgia AC	Georgia Asphalt Cement Price Index, English, Unit:$/Ton
NHCCI	National Highway Construction Index
Crude Oil Price	National level crude oil price
PPI: Diesel	Producer Price Index: No. 2 diesel fuel, Fuels and related products and power
ENR Labor	ENR Common Labor Index
PPI: Machine	Producer Price Index Industry Data, Construction and MFG.

It is interested to know how these factors affect the long-term pattern of unit bidding price. The preliminary correlation tests were performed (both parametric correlation test, Pearson correlation test, and non-parametric Spearman correlation test) and results are demonstrated in the following table:

Table 4-3. Correlation Efficient and p Value

Indicators	Pearson Corr.	p value for Pearson	Spearman Corr.	p value for Spearman
Georgia AC	0.5428	0	0.529	0
NHCCI	0.2952	0	0.415	0
Crude Oil Price	0.3121	0	0.3457	0
PPI: Diesel	0.4249	0	0.3558	0
ENR Labor	0.3898	0	0.4781	0
PPI: Machine	0.4421	0	0.4767	0

The results show that even though all factors have a linear or rank relation to bidding price (all p value is really close to 0) based on Pearson test and Spearman test respectively. However, the correlation coefficients are in non-significant level: only one factor (Georgia AC) is bigger than 0.5. This phenomenon is due to the fact that even though in a long run, the change trend correlation is significant different from 0, but the linear and rank relation is too weak to model the bidding price with these factors. The problem reflected from the test mainly comes from the inner trend change of bidding price and the big volatility compared to national level index, which reduces the fluctuations of local index based on statistical law of large number. First, the change trend of bidding prices does not maintain the same from 2008 to 2016. Second, in a period with a change trend the big volatility of the bidding price weakens the linear and rank relation among the indicators.

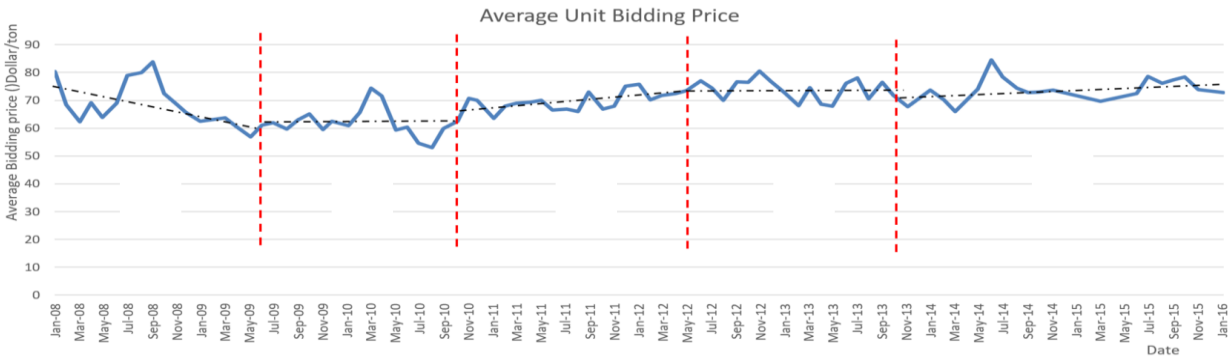


Figure 4-14. Five segments for bidding price after change point detection

Four change points are detected which separated the whole data set into five pieces, each of them are labeled as a character from “a” to “e”. The statistics for each piece are summarized in the following table.

Four change points correspond to four months. The p value of MK test is the statistic to test if there is a monotonic trend and the null hypothesis is there is a monotonic trend significantly different from zero. Therefore, piece a, c and e have a monotonic trend and then Theil-Sen estimator helps them to find the slope. The p value of piece e is not less than 0.05 but 0.144 is acceptable and the problem comes from the fluctuation from Mar.14 to Jul.14, which will noted in the latter. Piece b and piece d do not have a trend significantly different from 0, but it could be noticed that the volatility in these two pieces is big. Even though the volatility is been recorded in the following table, this reserch will not do quantitative research on it.

Table 4-4. Statistics for Five Segments

	a	b	c	d	e
Change point	Jun.09	Oct.10	May.12	Oct.13	-
p value of MK	0.0134	0.3	0.03	0.65	0.144
Theil-Sen slope	-0.98	-	0.411	-	-
Confidence interval of the slope	(-2.21,-0.145)	-	(0.014,0.682)	-	-
Mean	68.42	62.04	69.22	74.02	73.85
Median	68.24	61.02	69.35	74.48	73.67
Description	Negative trend, big volatility	No trend, big volatility	Positive trend, small volatility	No trend, big volatility	No trend, big volatility

The next step is doing the statistical analysis on related indicators and then finding their relations. The following example is given in Georgia asphalt and cement index. The analysis is based on the time period segmentation from change point detection of bidding price. In each segmentation the trend test is performed to decide whether it is necessary to find the slope.

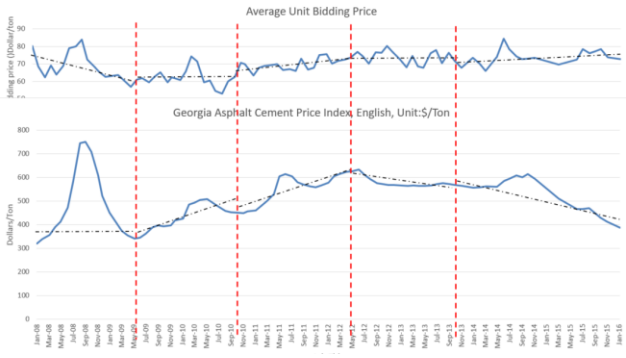


Figure 4-15. Comparative analysis of bidding price and AC index

The same analysis is done for all other five indicators and the result are shown in the following table, where “0” means no trend, “+” means a positive trend existed and “-”

means there is a negative trend. Based on the table, some possible deductions could be made about the relation among six chosen indicators and bidding price. The labor index seems to have few impacts on the bidding price because resurfacing projects are not man-power intensive work, instead, it more relies on machine. NHCCI seems to be the leading indicator that could be used to predict the long term trend of bidding price.

Table 4-5. Trend Analysis Result

Bidding Price	Asphalt Cement	Machine	NHCCI	Crude oil	Diesel	Labor
-	0	+	-	-	-	+
0	+	-	-	+	+	+
+	+	+	+	+	+	+
0	-	+	-	+	0	+
0	-	+	+	-	-	+

The proposed analysis framework could find more useful information than traditional Pearson and Spearman correlation test, such as how does the index go (upwards, downwards, no trend) in a period of time. There are two major limitations in this research: the first one is how to use the information of other indicators to predict the trend of bidding price. To solve this problem, more research will be emphasized on finding the appropriate lag between those indicators and bidding price. The second limitation is this research did not quantitatively consider the volatility. The trend and volatility should be the two components to comprehensively depict the long-term pattern of the bidding price. These two limitations form the future research plan to find out more information of bidding price.

4.4.2 Feature Selection

In this research, the authors used the Boruta importance analysis to implement the feature selection for the following two reasons: theoretically, it is one of the algorithms that express high computational speed when dealing with a large number of variables; it is useful when dealing with nonlinear related variables. The idea of Boruta is to find the variables that have the most information to make the prediction and rank them in order. In essence, the Borura algorithm is an ensemble method in which classification is performed by voting of multiple decision trees. These computed "votes" are used to rank the importance of the feature.

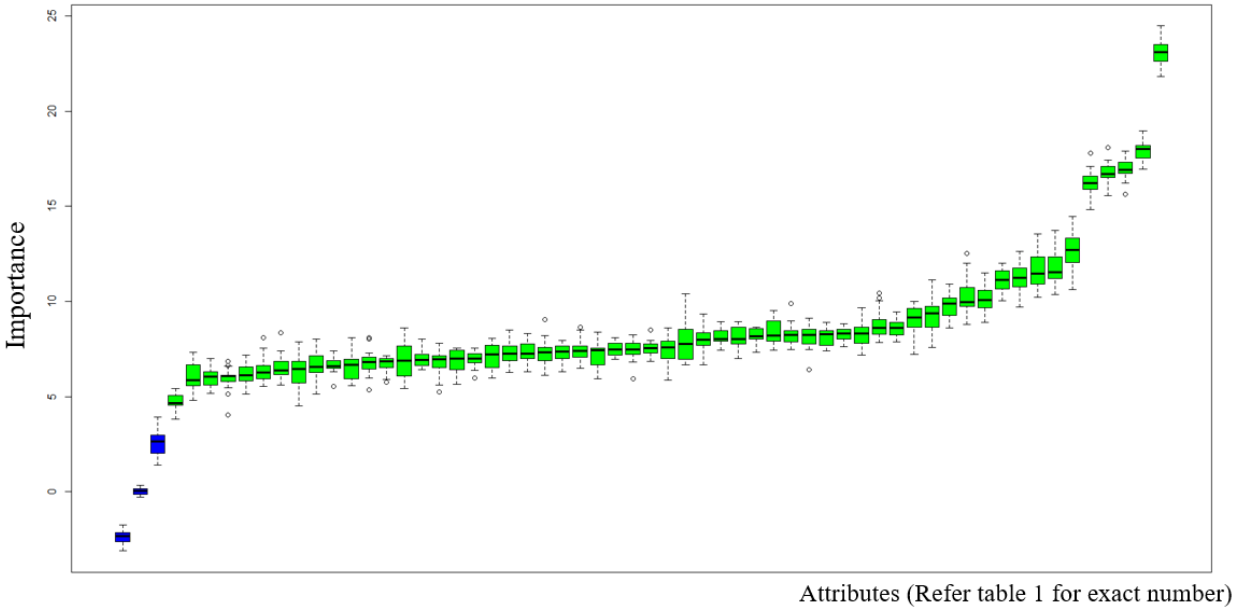


Figure 4-16. Running result of Boruta analysis

The result of implementing the Boruta analysis is displayed the figure. For the purpose of clearly displaying the figure, all features have been replaced by "x" plus a number. The required explanation will be given directly after each notation. The vertical coordination is the numerical calculation of feature importance. Boruta analysis marked x2

(project year) and x15 (last month county asphalt volume) as the "unimportant features" and they will not be used to train the model, while the rest of the features are "important." Some of the most important features that contain the most information to predict bidding price include x5 (terrain), x8 (project asphalt quantity), x6 (region number), x13 (number of asphalt plants within 80.5 kilometers), and x10 (project length).

Table 4-6. Boruta analysis results

Feature (Variable)	Index	Importance Measure
Month	x1	6.488
Year	x2	4.886
FIPS	x3	6.968
Primary County	x4	8.786
Terrain	x5	23.269
REGION	x6	17.386
Project Duration	x7	10.271
Quantity	x8	18.707
Total Bid Price	x9	12.043
Project Length	x10	15.934
Number of Bidders	x11	11.458
Number of Pay-items	x12	7.672
No. Asphalt Plants Within 50mi	x13	17.238
Monthly County Asphalt Volume	x14	11.084
Monthly County Number of projects	x15	5.747
Monthly GA Total Asphalt Volume	x16	6.367
Monthly GA Number of Total Contracts	x17	6.562
Monthly GA Total Contract Price	x18	6.509
Architecture Billings Index	x19	8.543
Building Permits for New Residential Construction	x20	8.157
Crude Oil Price	x21	7.375
DJIA	x22	10.314
ENR Building Cost Index	x23	7.470
ENR Common Labor Index	x24	6.644
ENR Construction Cost Index	x25	6.917
ENR Material Price Index	x26	8.209
ENR Skilled Labor Index	x27	7.163
Equipment Operator Wages, Paving, Mean hourly wage, Georgia	x28	6.881
GA Fuel Price Index	x29	8.111
GA Asphalt Cement Price Index	x30	8.171
GDP (GA Construction, in thousands)	x31	5.815

Table 4-6 continued

Inflation Rate	x32	6.701
JOLT, Hires (National, Construction, in Thousands)	x33	7.097
HMI (South)	x34	8.200
National Highway Construction Cost Index	x35	7.797
PPI, gasoline, fuels and related products and power	x36	7.522
PPI, steel mill products, metals and metal products	x37	8.258
PPI, No. 2 diesel fuel, fuels and related products and power	x38	6.411
PPI, crude petroleum (domestic production)	x39	7.581
PPI, construction machinery manufacturing	x40	7.389
Turner Cost Index	x41	8.163
Consumer Price Index (south)	x42	8.852
Diesel Retail Prices, Lower Atlantic	x43	6.762
Unemployment	x44	9.439
CPI (south)	x45	8.215
CPI (GA all construction)	x46	7.742
FMI NRCI	x47	6.896
Labor Productivity in Highway, Street, and Bridge Construction (National)	x48	6.528
PPI (Construction sand and gravel mining, National)	x49	7.463
Number of Establishments in private Construction Industry, County	x50	11.422
Average weekly wage, all industry (QCEW), County	x51	11.633
GA AC Price (12-month percentage change)	x52	7.115
GDP (levels) (12-month percentage change)	x53	8.401
JOLT, Hires (12-month percentage change)	x54	6.362
Unemployment (12-month percentage change)	x55	8.969
CPI (south, 12-month percentage change)	x56	6.386
Total dollar value of projects awarded in the same month at county level	x57	8.284

The first 20 “most important” features are selected to train the model and all features are displayed in the table 1 below. All the features that are not classified as “unimportant” could be used for model training, but researchers could select a subgroup of “important features” based on their knowledge or the performance of the model. In this research, the selection of the subset of important features is based on the recommended critical threshold for selecting important features in the Boruta feature selection algorithm,

which is 8.25 (Lin et al. 2015). Features with importance measures of 8.25 or lower are not selected by the Boruta algorithm in further development of the prediction model.

To further support the rationale of the selection of 20 most important features, a followed research was conducted by adding feature one by one based on their importance rank and plot the in-sample prediction error measured in MAPE. For example, when the number of features included is one, that means the model only used the rank-first feature, terrain type in predicting the unit price bids.

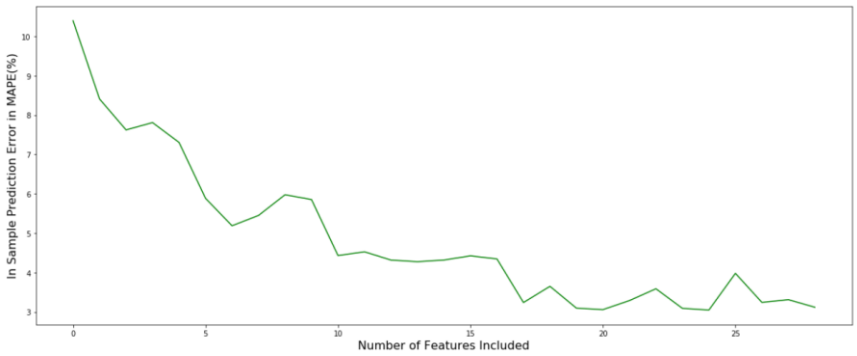


Figure 4-17. Plot the MAPE when the number of features is increased

Based on the above plot, the MAPE decreases significantly when the first 10 features were included. The following 10 features devoted another small improvement of the in-sample prediction accuracy. It was approximately estimated that when the number of features is bigger than 20, extra information could not help the model make better prediction.

Based on the above importance, the partial dependence plots are used for further analyzing the related factors. This research found the significant relation for the following four factors and some analysis is discussed.

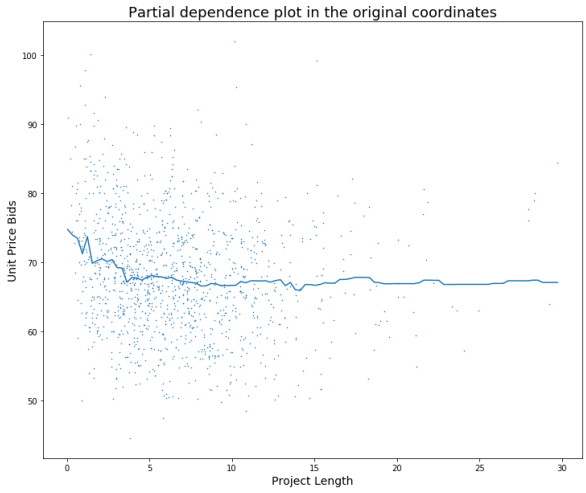


Figure 4-18. Partial dependent plot for project length

The unit price bids have a negative relation to the project length when the length of pavement is shorter than 5 miles. This works like the theory of economies of scale, which says the unit cost could be reduced when the total output of a product is increased. However, when project length is longer than 5 miles, there seems no significant correlation between them.

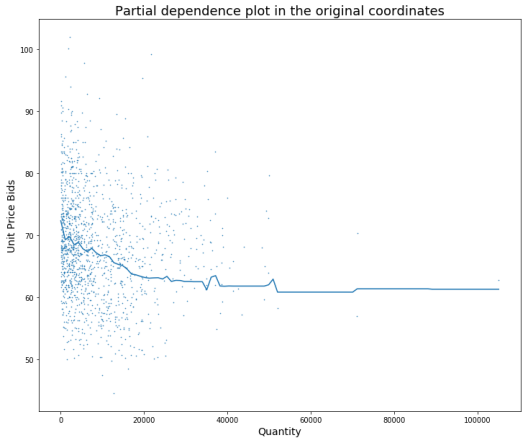


Figure 4-19. Partial dependent plot for quantity

A similar pattern was detected for the quantity. When the total quantity is smaller than 20000 tons, the unit price bids have a negative correlation with the quantity. There seems no significant relation between them when the quantity is larger than 20000. This

result is consistent with the last one because longer pavement length corresponds to the more consuming quantity of the asphalt line items.

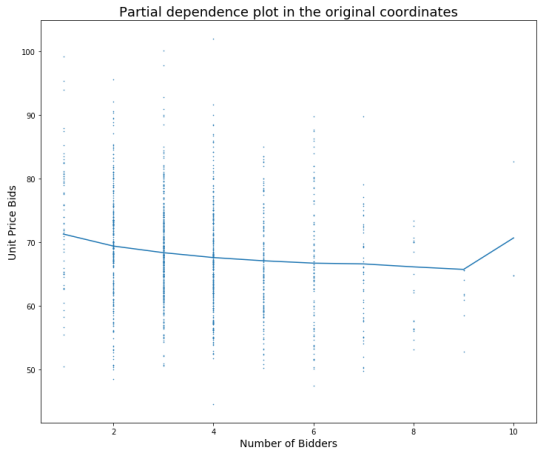


Figure 4-20. Partial dependent plot for the number of bidders

The third importance factor is the number of bidders. There is a negative relation between the unit prices bids and the number bidders. More bidders are helpful to decrease the price possible due to the higher competition. In this plot, the unit price bids suddenly go up when the number of bidders is 10. This maybe due to the small sample size of the data and in some sense those points could be determined as outliers.

The following one important factor is a categorical value: terrain. It is important means that there is significant difference among different categories.

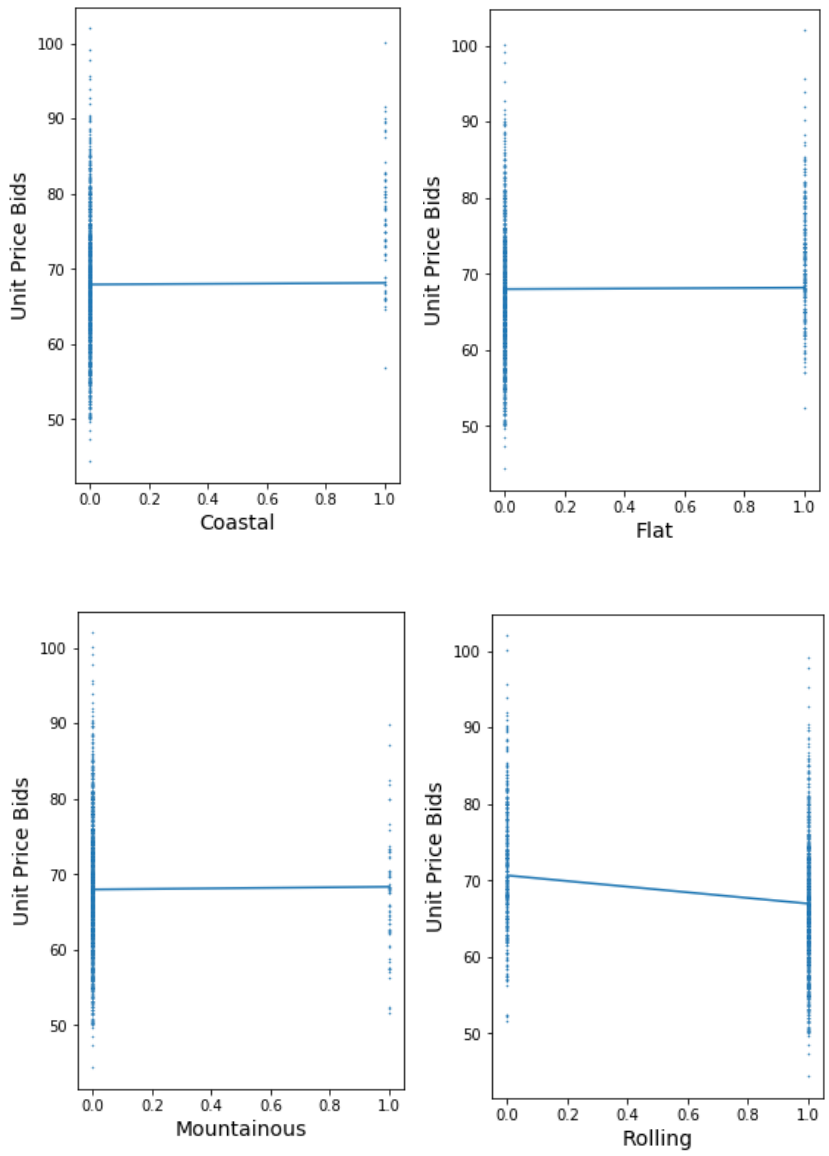


Figure 4-21. Partial dependent plot for terrain

With the figure, one thing needs to be mentioned is that it does not mean that the unit price bids are significantly different in each figure. As long as in at least one of the categories, if the price could exhibit a difference, then the feature could be a useful one. In this case, we see that the projects in rolling terrain have a lower cost than those not in this terrain.

4.4.3 Predicting the Unit Price Bids through Ensemble Learning

Data are separated into two parts, training data and testing data. Training data are used for model training and selecting the best parameters, while testing data are used for measuring the goodness of the model. The training set contains the data of projects ranging from January 2008 to December 2014, and the testing set uses the rest of the data which has 72 data points. In a machine learning research, all the data are randomly shuffled before dividing them into training and testing parts. In this research, this step will be ignored and all the data points will be kept in the time sequence, and the capability of the predictive model is examined using the roll forward cross validation approach. In the training set, there are a total of 1284 data points and each data point is a 22-dimensional vector, where there is one project index dimension, one bidding price dimension, and 20 variables (features).

In the process of model training in machine learning algorithms, the training set needs to be further divided into training data and cross-validation data. Unlike with the time series model fitting, where all the parameters could be estimated automatically, some parameters in the machine learning algorithms could not be automatically determined in training; for example, the number of layers and the number of nodes in each layer in the neural network should be decided before model training. These parameters are called the hyper-parameters. Cross-validation data are used to select the model with optimal hyper-parameters that could not be determined automatically. In this research, the five folds cross-validation is applied, which means that 80% of the training data will be used for training part of the parameters and the remaining 20% of the training data will be used for selecting hyper-parameters.

The trained model is tested by the test set, which has 101 data points, corresponding to the projects ranging from January 2015 to January 2016. This test set has never been used in the process of model training and cross-validation, so it could reflect the performance of the model to predict unknown and future events.

Three metrics will be used in this research to evaluate the model and compare the prediction accuracy:

Mean absolute error (MAE):

$$MAE = \frac{\sum_{i=1}^{n} |e_i|}{n} \quad 2.$$

Mean square error (MSE):

$$MSE = \frac{\sum_{i=1}^{n} |e_i|^2}{n} \quad 3.$$

Mean absolute percentage error:

$$MAPE = \frac{\sum_{i=1}^{n} |e_i|}{Y_{real} \times n} \times 100 \quad 4.$$

In these equations, e_i means the error derived from the difference between real value and predicted value. Results of three first-level models and the one second layer are displayed in the table for comparison, and some conclusions are provided here:

- For the three first-level models, extreme gradient boosting helps the researchers get the most accurate result, and random forest provides the worst result compared to the other two algorithms.

- The application of the ensemble learning model significantly improves the accuracy and amplifies the benefits of machine learning algorithms. The result of the ensemble learning model is much better than the three single ones in the first level. The benefit of the research results from the fact that even a tiny difference of the prediction accuracy on the unit price bids could impact a lot on the overall cost of a project.

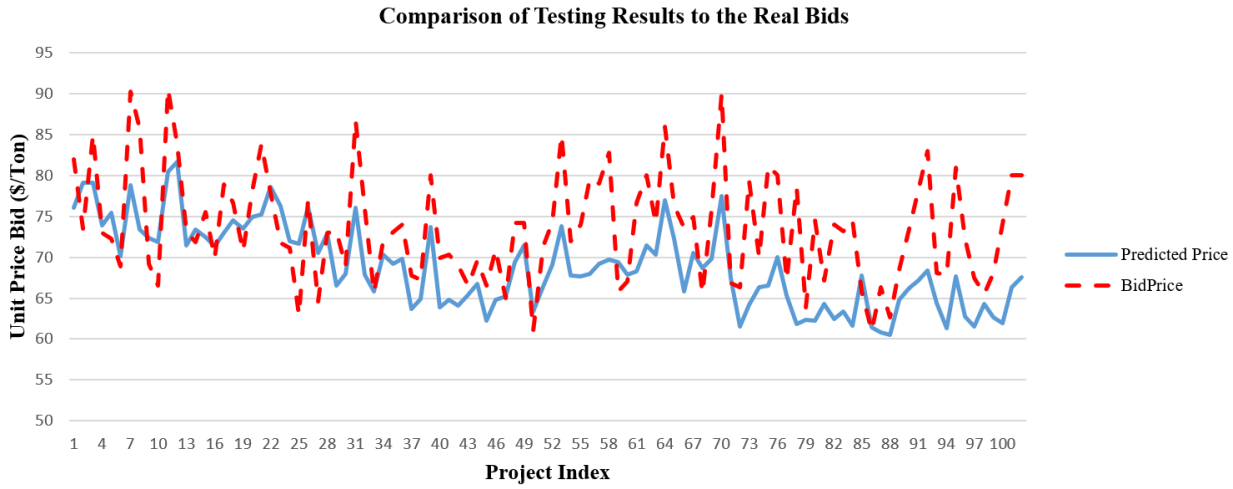


Figure 4-22. Testing result plot

Table 4-7. Test results

Error	gbm (gradient boosting)	xgbm (extreme gradient boosting)	rngr (random forest)	nnet (neural network)
MAE	7.7965	6.1104	8.2056	5.5412
MSE	99.2985	71.8564	114.2202	52.3049
MAPE	10.9854	7.6441	10.4716	7.5612

The comparison of the final prediction to the actual bidding price is plotted in Figure 5. This is not a time series because many projects may take place within the same month. By connecting the data points to a broken line, the researchers could see the model is able to roughly track the changing trend of the bidding price.

The first baseline model is based on the industry practice proposed in the literature from Back (Back et al. 2000), who applied the Monte Carlo simulation to modeling the construction cost data. The advantage of this method when compared to another NCHRP report (Anderson et al. 2006) is that it considers the uncertainty of construction cost, which should be considered in the unit price bid.

To implement this estimation method, the preliminary manipulation of the data makes the prediction problem to be a univariate time series question: the multiple data points in a same month are fused into one point by taking the average of them. In the training set, it provides us the information on the distribution of the average monthly cost escalation rates. With this information, Monte Carlo simulation helps to generate upcoming escalation rates and get the estimated future unit price bids. A disadvantage is that for all projects in month, it could only make a single prediction. Same to the data manipulation in the former part, the training set is from January 2008 to December 2014 and the rest of the data are the testing set.

Another baseline model is the multiple linear regression, which is one of the most prevailing prediction technics in highway cost research.

The matrix form of the regression model could be expressed as:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\epsilon}, \qquad 5.$$

where X is the design matrix that contains the data of 20 explanatory variables and one intercept; y is the vector representing the value of the unit price bids; βis the coefficients vector the researchers want to estimate; and ε is the error vector. Beta is the least-square estimator and the solution is derived from the following matrix equation:

$$\boldsymbol{b} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} \quad 6.$$

The goodness of fit is conducted by analysis of variance (ANOVA) and three measures are used for later calculation of statistics: total sum of squares (SST), sum of squares due to regression (SSR), and the sum of squared errors (SSE).

The R square in the regression is calculated as 0.47, which implies that the variables (features) are not significantly linearly related to the value of the unit price bids, but the model could be further evaluated to see its performance on the test set. If the new observations in the test set are denoted by X_t, the prediction is calculated through:

$$\boldsymbol{Y} = \boldsymbol{X_t} * \boldsymbol{b} \quad 7.$$

The results of MAE, MSE, MAPE, along with those in machine learning models, are shown in the following table for comparison.

Table 4-8. Test results compared with two benchmark models

Error	gbm	xgbm	rngr	nnet	Monte Carlo	multiple regression
MAE	7.7965	6.1104	8.2056	5.5412	7.539	8.888
MSE	99.2985	71.8564	114.2202	52.3049	102.936	117.102
MAPE	10.9854	7.6441	10.4716	7.5612	9.862	11.355

Every machine learning model works better than the multiple regression in this dataset. The ensemble learning method presented in this research provides a more accurate cost forecast compared to the existing methods. Using the mean absolute error (MAE), the ensemble learning method is 37.98% more accurate than the regression method, and 26.89% more accurate than the Monte Carlo simulation method. Similar improvements are reported in enhancing the cost forecasts based on other error measures, MSE and MAPE. From the statistics, the research team can conclude that compared to machine learning technics, the Monte Carlo simulation and the multiple regression model could not provide a better prediction accuracy for unit price bids. The application of Monte Carlo is limited to the capability on single prediction per month. The application of regression technique is limited to the linear relationship between variables (features) and the response. Machine learning algorithms are better at catching more complex, nonlinear relationships, and that is why they could solve more practical prediction problems.

According to the National Cooperative Highway Research Program (NCHRP) research report 08-49 "Procedures for estimation and management for highway projects during planning, programming, and preconstruction," conceptual estimating techniques with few project definitions can produce a project estimate with an accuracy range of +40/-20% to +120/-60%. Similar accuracy ranges are reported in the Association for the Advancement of Cost Engineering International (AACE) Total Cost Management Framework (Stephenson 2015). The researchers also reached out to subject matter experts in the Georgia DOT and they also confirmed that ±20-40% is an acceptable range of accuracy for forecasting construction cost at the early phase of project planning and conceptual design. The average estimation error using the new ensemble learning method

is approximately 7.56% measured in mean absolute percentage error, which is much more accurate than the desired accuracy level by GDOT and other transportation agencies.

4.5 Conclusions

Resurfacing projects is one of the most common highway projects in Georgia with a large amount of investment every year. The changing trend and big volatility of unit price bids make the problem in modeling and prediction be difficult. The research in the long-term trend analysis proposed a non-parametric framework to detect the change point in a robust way. The biggest advantage is the capability to work with any kinds of data in the industry. The non-parametric slope estimator is stable and robust. The application of non-linear feature selection methods fills the gap of using the linear Pearson correlation analysis. Non-linear relation is widespread in the industry and therefore describing this relation could make a huge improvement in modeling the highway construction cost. The ensemble learning method presented in this research provides a more accurate cost forecast compared to the existing methods. Using the mean absolute error (MAE), the ensemble learning method is 37.98% more accurate than the regression method, and 26.89% more accurate than the Monte Carlo simulation method. Similar improvements are reported in enhancing the cost forecasts based on other error measures, MSE and MAPE. In this research problem, a tiny improvement of the unit price bid suggests a huge advancement of the accuracy in budget estimation.

5. SUMMARY AND CONCLUSIONS

5.1 Summary and Contributions

The contributions to the body of knowledge in the field of highway construction cost research could be summarized in following aspects.

First, this research proposed a new model in time series prediction using LSTM. The proposed model in modified from the encoder and decoder architecture. It is not only efficient to the problem in the thesis, but also applicable to other numerical time series prediction problems.

Second, the proposed models contribute to a more accurate cost estimation: in terms of the out-of-sample accuracy, the LSTM model outperforms the seasonal ARIMA in three prediction scenarios. The ensemble learning method provides a more accurate cost forecast compared to the Monte Carlo simulation method and linear regression. In this field of research problem, a tiny improvement of the prediction suggests a huge advancement of the accuracy in budget estimation.

Third, from the practical point of view, all techniques selected in the research set no restrictions on the data. The proposed forecasting methods can handle a wide range of issues with the input data that are common in the field of highway construction cost forecasting. For example, the ensemble learning algorithm is a robust method that utilizes the extreme boosting algorithm that can handle missing values in the inputs data through using surrogate variables as substitutes for the predictors. The dataset used in the research contains several N/A (not applicable) as values for several features. The N/A's indicates

missing values for the selected features that can be due to missing records and human-related errors in collecting, entering and processing the data in practice. In the contrary, other forecasting methods, such as regression and time series analysis will get crashed before manually handle missing data, either by eliminating the entire data observation or by assuming or estimating the missing values. Either of these approaches can affect the accuracy of the forecasting results.

Also, the proposed models can efficiently handle both numerical and categorical variables. Not like linear regression, the categorical variables need to be treated as several dummy variables, or in some other methods the preliminary manipulation is required to manually transform categorical variables into numerical ones, all implemented algorithms in this research could automatically detect the categories with built-in converting function.

Fourth, this research emphasized the non-linear relation modeling and the explanatory power. The Boruta feature selection algorithm is a unique aspect of the proposed research method that facilitates the selection of the most relevant features with the greatest advantage for enhancing the cost prediction. Partial dependence plot helps engineers to easily understand the model as linear regression. The feature selection algorithm is particularly useful in the context of forecasting highway construction cost as it can select the most important features form a large number of variables to improve cost forecasting. Especially, the selection is challenging as the relationships among the features and the unit price bid are complex and nonlinear. The feature selection algorithm is shown to be a successful method to find a best subset of the features to develop a forecasting model with an outstanding performance as far as accurate prediction.

The proposed methods are computationally fast that makes it efficient to use in the real-world applications. The extreme boosting algorithm package provided by R has a built-in parallel computing algorithm that makes it significantly faster than other machine learning algorithms in dealing with the dataset of similar sizes to achieve a similar level of prediction accuracy. This is a great advantage of the proposed ensemble learning method that makes it a particularly attractive choice to handle larger datasets with more features.

In summary, the proposed forecasting models are ready for implementation. The models are applicable to different kinds of data in construction industry even with missing values. The prediction is stable and efficient compared to other models. With the improved prediction capability for HCCI and unit price bids for major line items, the proposed models provide great benefits to state DOTs in preparing more accurate budgets and cost estimates for highway projects.

5.2 Limitation and Future Research

Even though this research proposes an ensemble learning model that could predict the unit price bids of resurfacing projects in an acceptable accuracy level, considerable future work could be done to improve this research from the aspect of data collection, algorithm optimization, and feature selection.

In the feature collection part, for the time-related features, the one-month lag is created between the predictors and the response variable (unit price bids), which implies a limitation of the model: with the current information, the model is only able to look forward one month, even though the model could be applicable for one year. In future research,

some effort could be made to improve this time lag and extend the time effect of the prediction. The exploration of the feature is an unceasing process. For example, the range and variance of the submitted bids for one project are related to the unit price bids but this research was not able to collect this feature. The consideration of the temperature and climate condition is also related. Even though the terrain type and region indirectly consider the difference of temperature and climate, more direct factors such as precipitation, average temperature and humidity could be collected to directly explain the effect of temperature and climate. Another important factor is the application of price adjustment clauses (PAC). In Georgia, PAC is used to adjust the asphalt cement price in a monthly basis for those projects with a long duration. If the asphalt cement price is greater or less than the let price, DOT will reimburse or deduct the corresponding amount before the payment. Future research is supposed to add these features.

For the LSTM, the biggest shortage is that the small dataset is not applicable in the model. Simple data will cause the model to be overfitting. Even though the initiative of the LSTM is to solve the complex problems, this requirement needs to be noticed in practical application. LSTM is problematic in explanatory power: due to the complex unit structure, it is difficult to explain what factors determine the pattern in the research.

Some reasonable simplified assumptions have been made in this research. These assumptions might be the area that could be used to further improve the prediction accuracy. First, there might be a better alternative of machine learning algorithms out of the scope of these authors. Trying diverse algorithms and the ensemble structure might improve the prediction accuracy. Second, some part of the model could be optimized. For example, this research uses the one-hidden-layer neural network and nine hidden units on

it; it is unknown if more complicated neural network structure could improve the prediction accuracy. Finally, the “voting” weights for each algorithm in the first level are assumed to be the same in this research. The weights could be different and there are many prevailing methods such as Bayesian voting to improve the ensemble learning model.

Further research could also be done with the focus of feature selection. In this research, Boruta analysis is used to rank the features and leave out the “unimportant” features. The first 20 most “important” features are used to train the model, but it is unknown if another subset of the features could improve the prediction accuracy. For example, there is a phenomenon in linear regression called the enhancement effect, which means it is always possible to improve the prediction performance with a difference combination of features.

REFERENCES

Anderson, S. D., Molenaar, K. R., and Schexnayder, C. J. (2006). "Guidance for cost estimation and management for highway projects during planning, programming, and preconstruction." NCHRP Rep. No. 574, Transportation Research Board, Washington, DC.

Akintoye, A., Bowen, P., and Hardcastle, C. (1998). "Macro-economic leading indicators of construction contract prices." Construction Management & Economics, 16(2), 159- 175.

Aric Jenkins. (2017). "President Trump again called for $1 trillion on infrastructure—without many details." ⟨http://fortune.com/2017/02/28/trump-congress-address-infrastructure-investment/⟩ (Feb. 2017).

ARTBA, American Road and Transportation Builders Association. (2014). "Highways Needing Resurfacing or Reconstruction." ⟨ http://www.artba.org/Media/PDFs/6.03.2014-_ARTBA_Conditions.pdf⟩ (2016).

ARTBA, American Road and Transportation Builders Association. (2016). "Fixing America's surface transportation act: A comprehensive analysis." ⟨http://www.artba.org/wp-content/uploads/2014/03/FASTAct_Publication.pdf⟩ (2016).

Ashuri, B., and Lu, J. (2010). "Time series analysis of ENR construction cost index." Journal of Construction Engineering and Management, 136(11), 1227–1237.

Ashuri, B., and Shahandashti, S. M. (2012). "Quantifying the relationship between construction cost index (CCI) and macroeconomic factors in the United States." In 48th ASC Annual International Conference Proceedings.

Back, W. E., Boles, W. W., and Fry, G. T. (2000). “Defining triangular probability distributions from historical cost data.” Journal of Construction Engineering and Management, 10.1061/(ASCE)0733-9364(2000)126:1(29), 29–37.

Baek, M. (2018). “Quantitative Analysis For Modeling Uncertainty In Construction Costs Of Transportation Projects With External Factors.” PhD thesis, School of Building Construction, Georgia Institute of Technology, Atlanta, United States.

Baldi, P., & Sadowski, P. J. 2013. “Understanding dropout.” Advances in neural information processing systems (pp. 2814-2822).

Bishop, C. M. (2006). Pattern recognition and machine learning, Spring-Verlag New York, 1–58.

California Department of Transportation. 2018. “Price Index for Selected Highway Construction Items, First Quarter Ending March 31, 2018”. Accessed Jan 10, 2019. http://ppmoe.dot.ca.gov/hq/esc/oe/cost_index/historical_reports/CCI_1QTR_2018.pdf

Cao, Y., Ashuri, B., and Baek, M. 2018 "Prediction of Unit Price Bids of Resurfacing Highway Projects through Ensemble Machine Learning." Journal of Computing in Civil Engineering. 32(5): 04018043. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000788.

Chou, J. S., and Lin, C. (2012). “Predicting disputes in public–private partnership projects: Classification and ensemble models.” Journal of Computing in Civil Engineering, 27(1), 51–60.

Comprehensive R Archive Network. 2019. “Interpreting Predictive Models Using Partial Dependence Plots”. Accessed Feb 2019. https://cran.r-project.org/web/packages/datarobot/vignettes/PartialDependence.html

Contracts and Market Analysis Branch, Colorado Department of Transportation. 2012. “CDOT’s Methodology for Preparing the Colorado Construction Cost Index (CCI)”. Accessed Jan 10, 2019. https://www.codot.gov/business/eema/documents/2012/2012Q2CCI.pdf

Damnjanovic, I., Anderson, S., Wimsatt, A., Reinschmidt, K., & Pandit, D. (2009). "Evaluation of ways and procedures to reduce construction cost and increase competition" (No. FHWA/TX-008/ 0-6011-1), Texas Transportation Institute, Texas A&M University System.

Dayton, K.J., Macdonald, D., & Hammond, P. (2006). "Recent trends in highway construction cost." Journal of Construction Management, 7(13), 375–384.

Dietterich, T. G. (2000, June). "Ensemble methods in machine learning." In International Workshop on Multiple Classifier Systems, Springer Berlin Heidelberg, 1–15.

L. Ding, W. Fang, H. Luo, P.E.D. Love, B. Zhong, X. Ouyang. 2018. "A deep hybrid learning model to detect unsafe behavior: integrating convolution neural networks and long short-term memory" Autom. Constr., 86 (Supplement C) (2018), pp. 118-124

Fang, W., Ding, L., Luo, H., P.E.D. Love. 2018. "Falls from height: a computer vision-based approach for safety harness detection." Autom. Constr. 91 (2018) 53–61, https://doi.org/10.1016/j.autcon.2018.02.018.

Ervin, E. (2007). "How to Protect Profit as Materials Prices Rise." Puget Sound Business Journal(Seattle). <http://seattle.bizjournals.com/seattle/stories/2007/06/11/focus10.html>. (May 2016).

Federal Highway Administration. 2017. "National Highway Construction Cost Index (NHCCI) 2.0". Accessed Jan 10, 2019. https://www.fhwa.dot.gov/policy/otps/nhcci/desc.cfm.

Floy, J., Golden, K., and Mcmurry, R. (2013). Standard Specifications Construction of Transportation Systems. Georgia Department of Transportation

Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. Annals of statistics, 1189-1232.

Gallagher, J. (2008). “Material price escalation: Managing the risks.” Construction eNewsletter.<http://www.troutmansanders.com/material-price-escalation-managing-the-risks-01-16-2008/>. May 2017.

Goh, B. H., and Teo, H. P. (2000). “Forecasting construction industry demand, price and productivity in Singapore: The Box-Jenkins approach.” Construction Management and Economics, 18, 607–618.

Gransberg, D.D., Jeong, H.D., Karaca, I., and Gardner, B. (2017) “Top-down construction cost estimating model using an artificial neural network.” (No. FHWA/MT-17-007/8227-001), Iowa State University, Institute for Transportation.

Huntsman, B., Glover, B., Huseynov, S., Wang, T., and Kuzio, J. 2018. “Highway Cost Index Estimator Tool”. Accessed Jan 15, 2019. https://static.tti.tamu.edu/tti.tamu.edu/documents/PRC-17-73.pdf.

Hwang, S. (2009). “Dynamic regression models for prediction of construction costs.” Journal of Construction Engineering and Management, 135(5), 360–367.

Ilbeigi, M., Ashuri, B., and Hui, Y. (2014, May). “A Stochastic Process to Model the Fluctuations of Asphalt Cement Price.” In Construction Research Congress, 1111–1118.

Ilbeigi, M., Ashuri, B., and Joukar, A. (2016). “Time-series analysis for forecasting asphalt- cement price.” Journal of Management in Engineering, 04016030.

Illinois DOT. (2016). “FY 2017–2022 proposed highway improvement program,” Springfield. ⟨http://www.idot.illinois.gov/Assets/uploads/files/Transportation-System/Reports/OP&P/HIP/2017-2022/FY17-22%20Executive%20Summary.pdf⟩ (Spring 2016).

Jeong, H. D., Gransberg, D. D., and Shrestha, K. J. 2017. “Advanced Methodology to Determine highway Construction Cost Index (HCCI)”. No. FHWA/MT-17-006/8232-001. Ames, IA: Iowa State Univ.

Joseph Shrestha, K., Jeong, H. D., & Gransberg, D. D. (2016). Current practices of highway construction cost index calculation and utilization. In Construction Research Congress 2016 (pp. 351-360).

Joukar, A., and Nahmens, I. 2015. "Volatility Forecast of Construction Cost Index Using General Autoregressive Conditional Heteroskedastic Method." Journal of Construction Engineering and Management 142 (2015), 1943-7862.

Kohavi, R., and John, G. H. (1997). "Wrappers for feature subset selection." Artificial Intelligence, 97, 273–324.

Lindquist, K., and M. Wendt. 2007. "Inflation Estimation Models: Synthesis Prepared for Aaron Butters, Systems Analysis & Program Development Manager." Accessed March 15, 2017. http://www.wsdot.wa.gov/NR/rdonlyres/95B41678-B868-4899-A8CD-3F2FAA3604BD/0/InflationEstimationModelsSynthesis51807FINAL.pdf.

Li, W., Hu, J., Zhang, Z., and Zhang, Y. 2018. "A Novel Traffic Flow Data Imputation Method for Traffic State Identification and Prediction Based on Spatio-Temporal Transportation Big Data".

Lin, D., Xu, X., and Pu, F. (2015). "Bayesian Information Criterion Based Feature Filtering for the Fusion of Multiple Features in High-Spatial-Resolution Satellite Scene Classification." Journal of Sensors, vol. 2015, Article ID 142612, 10 pages, 2015. doi:10.1155/2015/142612.

Lowe, D. J., Emsley, M. W., and Harding, A. (2006). "Predicting construction cost using multiple regression techniques." Journal of Construction Engineering and Management, 132(7), 750–758.

Minnesota Department of Transportation. 2018. "Highway Construction Costs and Cost Inflation Study". Accessed Jan 10, 2019. https://www.dot.state.mn.us/govrel/reports/2018/2018-hwy-const-costs-and-cost-inflation-study.pdf.

Minnesota Department of Transportation. 2017. “Highway Construction Cost Index”. Accessed Jan 10, 2019. http://www.dot.state.mn.us/bidlet/CostIndex/CostIndexQ32016.pdf.

Moon, S., Chi, S., and Kim, D. Y. 2018 “Predicting Construction Cost Index Using the Autoregressive Fractionally Integrated Moving Average Model,” Journal of Management in Engineering, vol. 34, no. 2, 2018.

Nassereddine, H., Whited, G. C., & Hanna, A. S. (2016). Developing a Chained Fisher Construction Cost Index for a State Highway Agency. Transportation Research Record, 2573(1), 149–156. https://doi.org/10.3141/2573-18

Office of Contracts, Iowa Department of Transportation. 2018. “Price Trend Index for Iowa Highway Construction”. Accessed Jan 10, 2019. https://www.iowadot.gov/contracts/lettings/PriceTrendIndex.pdf.

Office of Estimating, Ohio Department of Transportation. 2013. “The Chained Fisher ODOT Construction Cost Index”. Accessed Jan 10, 2019. http://www.dot.state.oh.us/Divisions/ConstructionMgt/Estimating/ODOT%20ChainedFisher%20CCI/Understanding%20the%20ODOT%20Chained-Fisher%20CCI.pdf

Sepp Hochreiter; Jürgen Schmidhuber. 1997. "Long short-term memory". Neural Computation. 9 (8): 1735–1780. doi:10.1162/neco.1997.9.8.1735. PMID 9377276.

Shane, J. S., Molenaar, K. R., Anderson, S., and Schexnayder, C. (2009). “Construction project cost escalation factors.” Journal of Management in Engineering, American Society of Civil Engineers, 25(4), 221–229.

Shahandashti, S. M. (2014, May). “Analysis of the temporal relationships between highway construction cost and indicators representing macroeconomic, and construction and energy market conditions.” In Construction Research Congress 2014: Construction in a Global Network, 1103–1110.

Shahandashti, S. M., and Ashuri, B. (2015). “Highway Construction Cost Forecasting UsingVector Error Correction Models.” Journal of Management in Engineering, 32(2), 04015040.

Shrestha, K. J., Jeong, H. D., and Gransberg, D. D. (2016). “Current Practices of Highway Construction Cost Index Calculation and Utilization.” Construction Research Congress 2016, 351–360.

Stephenson, H. L.. (2017). “Total Cost Management Framework, an Integrated Approach to Portfolio, Program, and Project Management.” AACE International.

Sutskever, I. Vinyals, O. & Le. Q. V. 2014. “Sequence to sequence learning with neural networks.” In Proc. Advances in Neural Information Processing Systems 27: 3104–3112.

Sundermeyer, M., Schlüter, R., & Ney, H. 2012. “LSTM neural networks for language modeling.” In Thirteenth annual conference of the international speech communication association, 194-197.

Texas Department of Transportation. 2019. “Highway Construction Index Report (2012 Base)”. Accessed Jan 10, 2019. ftp://ftp.dot.state.tx.us/pub/txdot-info/cst/hci-binder.pdf.

Thomas Ng, S., Cheung, S. O., Martin Skitmore, R., Lam, K. C., and Wong, L. Y. (2000). “Prediction of tender price index directional changes.” Construction Management and Economics, 18(7), 843–852.

Utah Department of Transportation. 2019. “Construction Cost Index Report”. Accessed Jan 10, 2019. https://www.udot.utah.gov/main/uconowner.gf?n=40328213719990945

U.S. Dept. of Transportation (U.S. DOT). (2007). “Highway statistics 2007.” Washington, DC.

Virginia Dept. of Transportation (VDOT). (2010). “VDOT annual budget (Fiscal year 2010–2011).” ⟨http://www.virginiadot.org/projects/reportsbudget. asp⟩ (Jan. 4, 2011).

Wang, J., and Ashuri, B. (2016). “Predicting ENR construction cost index using machine learning algorithms.” International Journal of Construction Education and Research, 1–17.

Wang, Y., and Liu, M. (2012). “Prices of highway resurfacing projects in economic downturn: Lessons learned and strategies forward.” Journal of Management in Engineering, 28(4), 391–397.

Washington State Department of Transportation. 2016. “WSDOT Highway Construction Costs”. Accessed Jan 10, 2019. https://www.wsdot.wa.gov/NR/rdonlyres/A8EE6CB0-46F6-4EE8-95A3-62E9B793F31C/0/CostIndexData.pdf

Williams, T. P. (1994). “Predicting changes in construction cost indexes using neural networks.” Journal of Construction Engineering and Management, 120(2), 306–320.

Williams, T. P., and Gong, J. (2014). “Predicting construction cost overruns using text mining, numerical data and ensemble classifiers.” Automation in Construction, 43, 23–29.

Wilmot, C. G., and Cheng, G. (2003). “Estimating future highway construction costs.” Journal of Construction Engineering and Management, 10.1061/(ASCE)0733-9364(2003)129:3(272), 272–279.

Wilmot, C. G., and Mei, B. (2005). “Neural network modeling of highway construction costs.” Journal of Construction Engineering and Management, 10.1061/(ASCE)0733-9364(2005)131:7(765), 765–771.

Zaremba, W., Sutskever, I., & Vinyals, O. 2014. “Recurrent neural network regularization.” arXiv preprint arXiv:1409.2329.

Z. Zhang; Y. Wang; P. Chen; and G. Yu. 2018. “Application of Long Short-Term Memory Neural Network for Multi-Step Travel Time Forecasting on Urban Expressways”. CICTP 2017: Transportation Reform and Change—Equity, Inclusiveness, Sharing, and Innovation.